

**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**

**FACULTAD DE CIENCIAS BIOLÓGICAS**

**UNIDAD DE POSTGRADO**

**Caracterización molecular mediante marcadores ISSR  
de una colección de 50 árboles clonales e híbridos de  
cacao (*Theobroma cacao* L.) de la UNAS-Tingo María**

**TESIS**

para optar al grado académico de Magíster en Biología Molecular

**AUTOR**

Julio Alfonso Chia Wong

**Lima-Perú**

**2009**

## **DEDICATORIA**

Dedico este manuscrito e investigación a los millones de agricultores quienes trabajan y velan por proporcionarnos alimento y un mundo mejor.

## **AGRADECIMIENTOS**

A todos los miembros de mi familia, a quienes les debo mucho. Por los ánimos y la permanente vocación de ayudarme en todo momento.

A mi asesora, la profesora Mery Suni, quien gracias a su paciencia y acertadas observaciones me permitieron terminar esta investigación. Y a los miembros de mi jurado, por sus certeras observaciones y su constante preocupación por mejorar la tesis.

A Dorita Aguilar, una gran amiga, no sé que hubiera hecho sin tu ayuda. Un especial agradecimiento a la Dra. Libertad Alzamora, quien siempre velo por ayudarme en los momentos difíciles.

A mis compañeros y amigos de la maestría: Nadia, Ronnie, Lidia, Michael, Susy, Analí, Johanna, Alejandro, por compartir tantas horas de estudio y conversaciones moleculares.

A mis excelentes profesores de la maestría: Pablo Ramírez, Débora Alvarado, Ruth García, Rina Ramírez, Jaime Descailleaux, Enrique Escobar, Cesar Puicón, Hilda Solís; quienes no solamente me impartieron los conocimientos, sino que como tutores, me guiaron en el importante camino hacia la especialización.

A todos mis amigos del CIP, donde pasé un grato período de aprendizaje con la Dra. Sylvie Priou, Liliam Gutarra, Eva Huaman, Fernando Calle, María, Heidy, Carmen, Erika, Betty, etc. y a los muchos becarios que siempre se encuentran en mi más alta estima.

Un especial agradecimiento a mis amigos y colegas: Eliana Yglesias (UNALM) y Luis García (UNAS) por el gran apoyo moral y el asesoramiento técnico.

Finalmente mis plegarias a Buda, por guiarme en el camino hacia la paz interior.

## **FINANCIAMIENTO**

Esta Tesis fue posible gracias al Financiamiento del Consejo Nacional de Ciencia Tecnología e Innovación Tecnológica (CONCYTEC). Esta institución apostó en mi y otros compañeros de la maestría dándonos la oportunidad de una Beca que me permitió desarrollar y concluir mis estudios y publicar la Tesis para la obtención del Grado de *Magister Scientiae*.

Asimismo, quisiera expresar un especial agradecimiento a la Srta. Erika Saavedra, del FONDECYT, quien con su paciencia y predisposición pude lograr terminar esta Tesis.

## CONTENIDO

	página
Agradecimientos.....	xii
Dedicatoria.....	xiii
Resumen.....	xiv
Abstract.....	xvi
 <b>1. INTRODUCCIÓN.....</b>	<b>1</b>
 <b>2. ANTECEDENTES</b>	
2.1 Características del cacao	
2.1.1. Hábitat y Agronomía.....	4
2.1.2. Botánica.....	4
2.2 Importancia del cacao.....	6
2.3 Taxonomía y Nomenclatura.....	10
2.4 Características de las principales razas de cacao	
2.4.1. Cacao Criollo.....	14
2.4.2. Cacao Forastero. .....	15
2.4.3. Cacao Trinitario. ....	16
2.4.4. Cacao “Nacional”.....	17
2.5 Centro de Origen.....	17
2.6 Recursos genéticos del cacao.....	19
2.7 Mejoramiento Genético. ....	21
2.8 Caracterización de los Recursos Genéticos.....	23
2.9 Caracterización con Marcadores Bioquímicos.....	25
2.10 Caracterización Molecular.....	26
2.11 Marcadores moleculares ISSR.	28

.....	
2.12	Taxonomía
	Numérica.....
	32
2.13	Estimación de la Semejanza Taxonómica mediante
	análisis
	multivariado.....
	33
<b>3.</b>	<b>HIPÓTESIS Y OBJETIVOS</b>
	35
<b>4.</b>	<b>MATERIALES Y MÉTODOS</b>
4.1	Ámbito de estudio.....
	36
4.2	Material vegetal.....
	37
4.3	Aislamiento y purificación del ADN.....
	38
4.4	Calidad del ADN.....
	39
4.5	Determinación de la Pureza y Concentración del ADN.....
	39
4.6	Pre-selección de iniciadores ISSR.....
	40
4.7	PCR con los 50 individuos del estudio.....
	40
4.8	Condiciones PCR-ISSR.....
	40
4.9	Electroforesis en gel denaturante de poliacrilamida (PAGE).....
	41
4.10	Colecta de los datos moleculares y determinación de
	valores
	descriptivos.....
	42
4.11	Análisis
	multivariado.....
	44
4.12	Análisis
	poblacional.....
	46
<b>5.</b>	<b>RESULTADOS</b>
5.1.	Pruebas de aislamiento y purificación del ADN.....
	49
5.2.	Calidad del ADN.....
	49
5.3.	Elección de los iniciadores.....
	50
5.4.	Ensayos ISSR de la colección completa.....
	51

5.5.	Análisis Multivariado.	
5.5.1	Agrupamiento UPGMA.....	52
5.5.2	Ordenamiento.....	55
5.5.3	Estructura y variación Poblacional.....	60
<b>6.</b>	<b>DISCUSIÓN</b>	
6.1	Elección de los genotipos.....	65
6.2	Condiciones del ADN y la PCR.....	66
6.3	Valores descriptivos de los marcadores.....	67
6.4	Agrupamiento.....	69
6.5	Ordenamiento.....	76
6.6	Validación interna.....	78
6.7	Análisis Poblacional.....	79
6.7.1	Porcentaje de loci polimórficos.....	80
6.7.2	Índices de Shannon y Diversidad Genética de Nei.....	80
6.7.3	Coeficiente de diferenciación genética $G_{st}$ .....	81
6.7.4	AMOVA.....	82
<b>7.</b>	<b>CONCLUSIONES.....</b>	<b>84</b>
<b>8.</b>	<b>RECOMENDACIONES.....</b>	<b>85</b>
<b>9.</b>	<b>REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>86</b>

10. **ANEXOS**..... 95-118

## ÍNDICE DE TABLAS

	página
Tabla N° 1: Clasificación en secciones taxonómicas del género <i>Theobroma</i> y ubicación de la especie de cacao.	12
Tabla N° 2: Grupos de cacao y distribución geográfica (Lachenaud, 1997 y Motamayor <i>et al.</i> , 2002).	13
Tabla N° 3: Ventajas y desventajas de los marcadores ISSR.	32
Tabla N° 4: Lista de individuos de cacao utilizados para el estudio con marcadores ISSR. Parte A: Clones nacionales, Parte B: Clones internacionales, Parte C: Híbridos Bioversity/UNAS.	36, 37
Tabla N° 5: Índices de pureza y concentraciones de ADN obtenidas en lecturas con biofotómetro para 4 accesiones de cacao.	49
Tabla N° 6: Lista de los 13 Iniciadores ISSR que dieron buena respuesta en la amplificación de 4 muestras (por duplicado) de ADN de cacao.	50
Tabla N° 7: Lista de Iniciadores ISSR utilizados para el estudio molecular de las 50 accesiones de cacao. El índice ISSR es la suma de todos los PIC de los loci para cada iniciador. La cantidad de fragmentos raros ( $p \leq 0,05$ ) es por genotipo encontrado.	51
Tabla N° 8: Cantidad de loci polimórficos y porcentaje correspondiente para cada población y subpoblación analizadas.	60
Tabla N° 9: Índices de información Shannon y Diversidad génica de Nei para cada población y subpoblación analizadas.	61
Tabla N° 10: Diferenciación genética poblacional respecto a todos los loci ( $G_{st}$ ).	61
Tabla N° 11: Matriz de Identidades Genéticas y Distancias Genéticas normalizadas (unbiased) de Nei.	62
Tabla N° 12: Análisis Molecular de Variancia con tres componentes: población, subpoblación y dentro de subpoblación.	64



## ÍNDICE DE FIGURAS

	página
Figura N° 1: Fotografía de un Híbrido Bioversity/UNAS I16 de cacao mostrando frutos inmaduros. E.E.A. Tulumayo – Tingo Maria.	6
Figura N° 2: Producción de cacao en el Perú por departamentos, año 2006 (DGPA, 2007).	8
Figura N° 3: Mapa de Latinoamérica mostrando la distribución postulada para los diversos grupos de cacao (elaborado a partir de Lachenaud, 1997 y Motamayor <i>et al.</i> , 2002)	15
Figura N° 4: Esquema de las cuatro formas principales de fruto del cacao Forastero (García, 2000).	16
Figura N° 5: Esquema mostrando el alineamiento de los iniciadores ISSR no anclados con respecto a un ADN molde (elaboración propia). Los <i>primers</i> ISSR anclados se muestran en color rojo y azul. Los casos se explican en el texto.	29
Figura N° 6: Esquema mostrando el alineamiento de iniciadores ISSR anclados con respecto a un ADN molde (tomado y modificado de: De Vicente y Fulton, 2003). Los <i>primers</i> ISSR anclados se muestran en color rojo y azul. Los casos se explican en el texto.	30
Figura N° 7: Esquema del programa PCR-ISSR en el termociclador <i>Eppendorf Mastercycle gradient 96</i> .	41
Figura N° 8: Fotografía de una cámara vertical de electroforesis en poliacrilamida construida artesanalmente. Se muestra junto con la fuente de poder.	42
Figura N° 9: Electroforesis en gel de agarosa al 1% mostrando la calidad de ADN aislado por duplicado para 4 accesiones de cacao. Se indica un control en el primer carril.	50
Figura N° 10: Electroforesis en gel de poliacrilamida al 6% mostrando los patrones de amplificación de 5 <i>primers</i> ISSR usados (UBC834, UBC835, UBC841, UBC842 y UBC884) con los genotipos cuyos códigos se muestran en el lado superior del carril. Marcador de peso molecular en ambos extremos: O'GeneRuler 100 bp DNA Ladder Plus (Fermentas).	52

Figura N° 11:	Fenograma generado con el coeficiente de asociación Jaccard. A un valor de 0,79 de similitud, se muestran 13 <i>clusters</i> en diferentes colores. El coeficiente de correlación cofenética (r) es de 0,91383.	53
Figura N° 12:	Fenograma de los híbridos y sus respectivos parentales generado con el coeficiente de asociación Jaccard. Los híbridos están resaltados con letra mayor. El coeficiente de correlación cofenética (r) es de 0,77013.	54
Figura N° 13:	Análisis de Coordenadas Principales, coeficiente Jaccard. Con el fin de una mejor visualización, sólo se muestran 46 OTUs (se excluyen UF29, H60, H37 y Seño). Se han encerrado los genotipos más afines entre sí.	56
Figura N° 14:	Escalamiento Multidimensional no métrico. Con el fin de una mejor visualización, sólo se muestran 46 OTUs (se excluyen UF29, H60, H37 y Seño). Se han encerrado los genotipos más afines entre sí. Valor de <i>Final STRESS2</i> = 0,10873.	57
Figura N° 15:	Análisis de Coordenadas Principales de los Híbridos Bioversity/ UNAS (recuadros) y sus respectivos parentales (elipses), coeficiente Jaccard, se señala con una flecha la relación con su progenie.	58
Figura N° 16:	Escalamiento Multidimensional no métrico de los Híbridos (recuadros) y sus respectivos parentales (círculos), coeficiente Jaccard, se señala con una flecha la relación con su progenie. Valor de <i>Final STRESS2</i> = 0,26212.	59
Figura N° 17:	Dendograma sin raíz de las subpoblaciones analizadas. Escala en la parte inferior izquierda. Se exportó el archivo del POPGENE al Treeview para su mejor visualización.	63

## ÍNDICE DE ANEXOS

	página
Anexo N° 1: Datos morfológicos más resaltantes recopilados de los clones e híbridos de cacao analizados. N.D.: no determinado.	95-97
Anexo N° 2: Lista de algunos caracteres agronómicos de importancia recopilados de los clones e híbridos de cacao analizados. NUSE: número de semillas, PESE: peso de semilla. N.D.: no determinado.	98,99
Anexo N° 3: Datos de Pureza ( $A_{260}/A_{280}$ ) y concentraciones de ADN (ng/μl) para las 50 accesiones de cacao. Lecturas obtenidas del biofotómetro Eppendorf.	100,101
Anexo N° 4: Lista de los 59 Iniciadores ISSR. Los códigos corresponden a iniciadores de la Universidad de British Columbia. Los códigos que resaltados en negrita pertenecen a los 13 iniciadores cuyos perfiles de amplificación fueron más consistentes que del resto.	102
Anexo N° 5: Datos Binarios (presencia: 1 ó ausencia: 0) de 73 loci (caracteres) para 50 genotipos (OTUs). El código 2 es el asignado a los datos perdidos o " <i>missing data</i> ". Se muestran sólo los loci con un porcentaje de datos perdidos menor o igual al 10%.	103-105
Anexo N° 6: Valores descriptivos calculados para cada locus. Loci con un porcentaje de datos perdidos $\leq 10\%$ . Mono = locus monomórfico.	106,107
Anexo N° 7: Correspondencia entre los 5 iniciadores utilizados y los genotipos que muestran fragmentos amplificados con una frecuencia alélica ( $p$ ) menor a 0,05 (indicados en el Anexo N° 6).	108
Anexo N° 8: Archivos Lote o " <i>Batch files</i> " del programa NTSYSpc v2.1p usados para el análisis multivariado de los datos moleculares.	109

Anexo N° 9:	Matriz de similitud de la comparación uno a uno entre las 50 OTUs. Coeficiente de Asociación Jaccard.	110-112
Anexo N° 10:	Fenograma UPGMA generado con los 50 individuos analizados, utilizando el coeficiente de similitud Dice.	113
Anexo N° 11:	Gráficos de la comparación de las dos matrices: cofenética y de distancia. Se presentan para el análisis de los 50 genotipos y para el análisis de los 6 híbridos + 11 parentales. Se muestran los coeficientes de correlación cofenética para cada gráfico.	114,115
Anexo N° 12:	Gráficos de ordenación en dos dimensiones representando a todos los individuos de la colección en Análisis de Coordenadas Principales y Escalamiento Multidimensional no métrico. Se puede observar la gran distancia que hay entre el grupo de 4 accesiones (UF29, H60, H37 y Seño) y el restante. En las Figuras N° 13 y 14 se representaron el resto de la colección.	116,117
Anexo N° 13:	Valores $G_{st}$ de la comparación par a par entre las subpoblaciones. Entre paréntesis se muestran el número de individuos por población.	118

## ABREVIATURAS

ADN	Acido desoxirribonucleico
AFLP	<i>Amplified Fragment Length Polymorphism</i> (polimorfismo de la longitud de los fragmentos amplificados)
AMOVA	Análisis Molecular de Variancia
ARN	Ácido ribonucleico
CATIE	Centro Agronómico Tropical de Investigación y Enseñanza
CIRAD	<i>Centre de coopération Internationale en Recherche Agronomique pour le Développement</i>
com. per.	Comunicación personal
CTAB	<i>Cetyl trimethyl ammonium bromide</i>
DGPA	Dirección General de Promoción Agraria
dNTP	Mezcla de desoxirribonucleótidos (dATP, dCTP, dGTP y dTTP)
EDTA	<i>Etilen Diamino Tetra Acetic acid</i> (ácido etilendiamino tetraacético)
E.E.A.	Estación Experimental Agrícola
$F_{ST}$	Diferenciación génica entre todas las subpoblaciones.
$G_{st}$	Diferenciación genética poblacional
ha (has)	Hectárea (hectáreas)
IBPGR	<i>International Board for Plant Genetic Resources</i>
ICCO	<i>International Cocoa Organization</i>
ICG,T	<i>International Cocoa Genebank at Trinidad</i>
IPGRI	<i>International Plant Genetic Resources Institute</i>
ISSR	<i>Interspread Single Sequence Repeats</i> (secuencias intercaladas entre microsatelites)
kb	Kilobases (mil pares de bases)
MAS	<i>Marker Assisted Selection</i>
MBD	matriz básica de datos
MINAG	Ministerio de Agricultura (Perú)
mg	Miligramos
m.s.n.m.	Metros sobre el nivel del mar

N.D.	No determinado
NEB	<i>Neutral Electrophoresis Buffer</i>
ng	Nanogramo
nmMDS	<i>Non metric Multidimensional Scaling</i> (análisis multidimensional no métrico)
OTUs	<i>Operational Taxonomic Units</i> (unidades taxonómicas operativas)
pb	pares de bases
PAGE	<i>Polyacrilamide Gel Electrophoresis</i> (electroforesis en gel de poliacrilamida)
PCR	<i>Polymerase Chain Reaction</i> (reacción en cadena de la Polimerasa)
PcoA	<i>Principal Coordinate Analysis</i> (análisis de coordenadas principales)
PIC	<i>Polymorphism Index Content</i> (índice de contenido polimórfico)
QTL	<i>Quantitative Trait Loci</i> (loci de caracteres cuantitativos)
RAPD	<i>Randomly Amplified Polimorphism DNA</i> (polimorfismo en el ADN amplificado al azar)
RFLP	<i>Restriction Fragment Length Polimorphism</i> (polimorfismo en la longitud de los fragmentos de restricción)
r.p.m.	revoluciones por minuto
SAHN	Sequential Agglomerative Hierarchical Nested cluster analysis
UBC	University of British Columbia
USDA	<i>United States Department of Agriculture</i>
SSR	<i>Single Sequence Repeats</i> (secuencias simples repetitivas o microsatélites)
TBE	<i>Tris Borate EDTA buffer</i> (tampón de electroforesis Tris Borato EDTA)
TE	tampón Tris-EDTA
TM	Toneladas métricas
µg	Microgramo
µl	Microlitro
UPGMA	<i>Unweighted Pair Group Method using Arithmetic Averages</i>
VNTR	<i>Variable Number Tandem Repeats</i> o minisatélites

## RESUMEN

Los programas de mejoramiento genético requieren de herramientas versátiles y de bajo costo que permitan lograr en menor tiempo nuevas variedades. En el caso del cacao, se disponen de una serie de técnicas moleculares que facilitan el trabajo al fitomejorador. Sin embargo, la utilidad y conveniencia de tales, serán elegibles dependiendo de las condiciones y objetivos del trabajo. Por este motivo, se trazaron los objetivos de ensayar los marcadores ISSR, determinar su poder discriminante y determinar parámetros de diversidad genética en una colección de cacao de Tingo María. Ésta consistió en 50 individuos o accesiones, de diversas procedencias genéticas y geográficas. El ISSR es una técnica basada en PCR que amplifica el fragmento entre dos secuencias de microsatélites, utilizando un iniciador complementario a éstos. Las muestras fueron colectadas en la provincia de Leoncio Prado (región de Huánuco) donde los árboles de cacao se mantienen en dos ubicaciones: el Banco de Germoplasma de la Universidad Nacional Agraria de la Selva y de la Estación Experimental Agrícola Tulumayo. Según el origen geográfico, se separaron 44 accesiones en dos poblaciones y once subpoblaciones. Las restantes seis accesiones no fueron incluidos en el análisis poblacional dado su origen híbrido. Se colectaron hojas sanas de uno o dos árboles clonales y se almacenaron en silicagel. Se determinaron preliminarmente los protocolos adecuados de extracción de ADN, PCR, y electroforesis con 4 accesiones: ICS1, CAT4, U70 y EET223; se experimentaron con estos genotipos con 59 iniciadores ISSR disponibles en las mismas condiciones PCR. Resultaron 13 *primers* ISSR con un perfil de amplificación aceptable, de los cuales se eligieron los 5 mejores para amplificar las 50 muestras. El análisis de datos se hizo con dos grupos de datos: un grupo abarcó los 50 OTUs totales y el otro grupo consistió en 6 híbridos Bioversity/UNAS con sus 11 parentales (total: 17). El análisis multivariado y genético se realizó utilizando tres paquetes estadísticos: NTSYSpc v2.1p, Arlequín v3.1 y POPGENE v1.32. Con el primer programa, se realizaron los análisis de agrupamiento (UPGMA-Jaccard) y ordenación (PCoA y nmMDS), así como los de validación (prueba de Mantel), logrando una diferenciación de cacao Nacionales vs. Internacionales, así como una distribución espacial entre parentales y su respectiva

progenie. Los estimadores genéticos y poblacionales se determinaron con los dos últimos. Los fenogramas y representaciones bidimensionales agruparon a las accesiones sin estricta relación geográfica ni morfológica, concordando con trabajos anteriores en diversidad molecular de cacao. Se estimaron los índices de diversidad de Shannon, los parámetros  $G_{st}$  y las distancias genéticas de Nei para las subpoblaciones, asumiendo un Equilibrio Hardy Weinberg, resultando que el cacao de Cusco y Huánuco poseen mayor diversidad genética que el resto. Asimismo, mediante el AMOVA se obtuvo que el nivel de variación a nivel intrasubpoblacional (76,8%) fue mayor que el intersubpoblacional (16,74%) y mucho mayor que el interpoblacional (6,46%). Se coincidió con diversos trabajos previos de diversidad genética de cacao con isoenzimas y RAPD (Parra *et al.*, 2001) y microsatélites (Zhang *et al.*, 2006; Motamayor *et al.*, 2008), en el sentido de que la variación intrasubpoblacional es considerablemente mayor que las demás variaciones. Entonces, según estos resultados, los marcadores ISSR podrían usarse como herramientas de bajo costo para facilitar el trabajo en Mejoramiento Genético de cacao.

**Palabras Clave:** *Theobroma cacao* L., ADN, ISSR, Caracterización molecular, Diversidad genética.



## **ABSTRACT**

In order to obtain new crop varieties in shorter time, plant breeding programs require versatile and inexpensive tools. For cacao crop, there are available a number of molecular techniques that support the breeder's job. Nonetheless, work objectives and laboratory conditions determine the utility and convenience of those techniques. In this research, the aim was to test the discriminatory power of ISSR molecular markers in a cacao collection in Tingo Maria, and to determine its genetic diversity. Fifty cacao accessions from different geographical and genetic origin constituted the sampled collection. The ISSR markers consist in DNA fragments amplified by PCR located between two adjacent microsatellite (SSR) loci with the same but inverted sequence. A single primer complementary to these microsatellite sequences is used. The leaf samples were collected from two locations at the Leoncio Prado province (Huanuco region): from the cacao germplasm bank of the Universidad Nacional Agraria de la Selva, and from the Agricultural Experimental Station "Tulumayo". A hierarchical grouping into 2 populations and 11 subpopulations of 44 accesions was performed taking into consideration the accession geographical origin. The remaining six accessions were excluded due to their hybrid origin. Leaves in good condition were sampled from one or two clonal trees. Then, they were kept and dried in silicagel. As a preliminary assay, DNA extraction and PCR amplification were carried out with four accessions: ICS1, CAT4, U70 and EET223. A group of 59 available ISSR primers were tested with these genotypes and 13 of them gave adequate amplification profiles. Then, five primers were selected and used to amplify the 50 accessions. Multivariate analyses were carried out separating 2 groups of data: the total 50 accessions constituted one group and the second group constituted by six Bioversity/UNAS hybrids and eleven parentals (total: 17 individuals). Three software packages were utilized: NTSYSpc v2.1p, Arlequin v3.1 and POPGENE v1.32. Grouping and ordination analysis, and internal validations were done with the first software. Genetic and population parameter estimation were calculated with the two latter programs. Accessions were distributed by phenograms and biplots with neither geographical nor morphoagronomical relationships, in agreement with previous cacao molecular diversity papers. Shannon index,  $G_{st}$  value, and Nei genetic distance were

calculated assuming Hardy-Weinberg equilibrium, finding that Cusco and Huanuco regions possess higher cacao genetic diversity. Furthermore, hierarchical AMOVA partitioned variation into three components: among populations 6,46%; among subpopulations within populations 16,74%; and within subpopulations 76,80%. This is in accordance with some previous genetic diversity research reports that found greater variation within subpopulation than the other ones (Parra *et al.*, 2001; Zhang *et al.*, 2006; Motamayor *et al.*, 2008). In conclusion with these results, it is advisable the utilization of ISSR in further cacao molecular characterization in national breeding programs.

**Keywords:** *Theobroma cacao* L., DNA, ISSR, Molecular characterization, Genetic diversity.

## 1. INTRODUCCIÓN

El cacao, *Theobroma cacao* L., es un cultivo tropical cuya importancia está en aumento principalmente para los agricultores pequeños de varias zonas tropicales de África del oeste, Centroamérica, Sudamérica y Asia. Esto se debe a que el mercado consumidor extiende cada vez más su demanda y diversidad de consumo, estimulando a que los países productores y consumidores inviertan en programas y estrategias de investigación con la finalidad de incrementar la productividad y la calidad del cacao y mejorar las condiciones de vida de los agricultores.

Con respecto al objetivo de aumentar la productividad del cultivo de cacao, dos factores son importantes: incrementar el rendimiento/ha y disminuir el costo de producción. Éstos pueden estar directamente relacionados al genotipo de la variedad, al control de plagas y enfermedades, a las técnicas de manejo agronómico, al mejoramiento del suelo y condiciones climáticas.

El mercado de los cacaos “finos” o de alta calidad está consolidándose cada vez más. Por esto, muchos fitomejoradores dirigen sus esfuerzos a aumentar la calidad del cacao como producto, implicando la selección de variedades con características excelentes en sabor y olor, así como la investigación y estandarización de las técnicas en poscosecha y fermentación.

Por último, pero no menos importante, en búsqueda del progreso del agricultor cacaotero, se deben vincular trabajo social, políticas adecuadas por parte del Estado y conseguir tratos justos de parte de las empresas privadas. Además, los centros de investigación deben extender de manera sostenida las innovaciones tecnológicas a los agricultores.

En el Perú, el cultivo del cacao es uno de los más importantes de la región tropical; junto con el café, la palma aceitera y los cítricos. Pese a su importancia, se obtienen rendimientos bajos y la calidad del producto es medianamente buena, por lo que aún no logra ser competitivo. En la década

de los 80's y 90's, el cacao estaba considerado como cultivo alternativo a la hoja de coca.

En consecuencia, muchos agricultores individuales y cooperativas están apostando por instalar nuevas plantaciones de cacao o renovar las antiguas con vista al futuro. Una de las principales limitaciones para este caso es que se disponen de pocas variedades con características ventajosas o cuyo producto tiene un buen precio en el mercado. Esto se debe en parte a que la obtención de variedades de cacao es un proceso lento debido al ciclo de vida largo que tiene la especie. Asimismo, a pesar que el Perú es uno de los centros de diversidad del cacao, las campañas de colecta y caracterización de germoplasma promisorio son escasas.

En la Universidad Nacional Agraria de la Selva (UNAS), entidad líder en la investigación agraria en la zona tropical del Perú, se está llevando a cabo un proyecto con fines de mejoramiento de cacao en convenio con Bioversity International llamado: "Mejoramiento de la Productividad y Calidad del Cacao: un enfoque participativo (CFC/ICCO/IPGRI) 2004-2009". Por medio de éste, se han logrado obtener diversas variedades híbridas promisorias de cacao con características importantes que satisfacerán la demanda local y regional, en términos de incremento de la producción y calidad del cacao. El material de partida para los trabajos de mejoramiento y selección fue posible gracias a la disponibilidad de un banco de germoplasma, establecido hace aproximadamente 20 años.

La continuidad de estos proyectos debe ser lograda a través de resultados importantes, con el objetivo de aminorar el tiempo y el costo de la obtención de las nuevas variedades. La tendencia en el mundo es el mejoramiento genético asistido con marcadores moleculares, estudiando los genes que permitirían acortar el tiempo y disminuir los costos que implica la selección de la progenie después de varios cruzamientos. Asimismo, la caracterización molecular de los materiales genéticos en los bancos de germoplasma tiene también un rol importante y, junto con los datos morfológicos y agronómicos, se obtendrán resultados más sólidos y confiables.

Los marcadores moleculares se basan en tres diferentes estrategias de detección del polimorfismo: hibridación del ADN, reacción en cadena de la polimerasa y secuenciamiento de ADN. Sin embargo, cada una de las mencionadas técnicas tiene ventajas y desventajas que deben evaluarse según la aplicación y condición (Karp *et al.*, 1997; Spooner *et al.*, 2005). En muchos casos, debido a la escasez de especialización técnica, recursos financieros e infraestructura adecuada el acceso a las técnicas en biología molecular de última generación es muy limitado y para esto se deben evaluar las alternativas disponibles.

A través de tesis de pregrado se han caracterizado gran parte de los materiales genéticos del banco de Germoplasma de la UNAS tanto a nivel morfológico, botánico como agronómico, y por medio de un proyecto colaborativo (Zhang *et al.*, 2006) se caracterizaron los materiales nacionales Ucayali y Huallaga mediante marcadores microsatélites. Sin embargo, este trabajo, así como otros (Motamayor *et al.*, 2008) de estudio molecular de cacao, demandaron infraestructura y costos inaccesibles para países en desarrollo, tal como un equipo de secuenciamiento de ADN con electroforesis capilar. Asimismo, el análisis molecular con microsatélites demanda la utilización de iniciadores o “*primers*” con un colorante fluorescente, cuyo precio es mayor que el de los iniciadores comunes.

En este programa de mejoramiento genético surgió la necesidad de caracterizar molecularmente los recursos genéticos ya disponibles y asimismo, evaluar la utilidad de los marcadores moleculares en la diferenciación del germoplasma en regiones de importante biodiversidad de cacao en Perú en condiciones locales. Para ello, se planteó la utilización de los marcadores semiarbitrarios: secuencias intercaladas entre microsatélites o *Interspread Single Sequence Repeats* (ISSR), cuyo costo es muy similar al de técnicas como el *Random Amplified Polimorphism DNA* (RAPD) pero con una comprobada mayor reproducibilidad (Charters y Wilkinson, 2000).

## **2. ANTECEDENTES**

### **2.1 Características del cacao**

#### **2.1.1 Hábitat y Agronomía**

Es un árbol generalmente de porte bajo, alcanzando alturas de 2 a 6 metros; sin embargo, ha llegado hasta los 25 metros en estado silvestre (Lachenaud *et al.*, 2007). Crece en el sotobosque ya que requiere de sombra, necesita protección del viento y un suelo rico y poroso. La altitud ideal para su desarrollo es de 400 m.s.n.m. Sin embargo, se le encuentra en altitudes que varían desde el nivel del mar hasta los 1000 m.s.n.m. El terreno debe ser rico en nitrógeno y en potasio. El clima debe ser húmedo tropical, una gran humedad y con temperatura que varía entre los 20 °C y los 30 °C, con una mínima de 16 °C (Braudeau, 1981).

A pesar de que sus frutos maduran durante todo el año, normalmente se realizan dos cosechas: la principal (que empieza hacia el final de la estación lluviosa y continúa hasta el inicio de la estación seca) y la intermedia (al principio del siguiente período de lluvias), y son necesarios de cinco a seis meses entre su fertilización y su recolección.

El cacao crece, se desarrolla y produce frutos en asociación con otras especies, que le proporcionan sombra y para conformar sistemas agroforestales sostenibles: cacao - gliricidia - pimienta; cacao - eritrina o cacao - cedro (Wood, 1975; Adriazola, 2003). Varios grupos indígenas cultivan el cacao en sus huertos familiares.

#### **2.1.2 Botánica**

El cacao posee una raíz principal pivotante; con hojas simples, enteras y de colores variables que van desde morado hasta verde pálido, con pecíolo corto; posee flores pequeñas, de color rosa, hermafrodita y pentámera, de

ovario súpero, con cinco lóculos donde hay de 6 a 12 óvulos/lóculo. Las flores al igual que los frutos se producen en racimos pequeños (llamados cojines florales), sobre el tejido maduro del tronco y de las ramas (conocido como cauliflora) (Braudeau, 1981).

Generalmente, su polinización es entomófila, llevada a cabo principalmente por varias especies de microinsectos alados del género *Forcipomyia* (Familia Ceratopogonidae, Orden Diptera). Una planta adulta puede llegar a producir más de 50 000 flores al año, de las cuales sólo se fecundan y llegar a fructificar entre el 0,5 y 2 % (Dias, 2001).

El fruto comúnmente conocido como “mazorca”, es de tipo baya indehisciente y madura entre 5 y 6 meses después de la polinización (Figura N° 1). Posee un mesocarpo de contextura lisa o arrugada que se divide en cinco carpelos interiormente. La rugosidad del fruto y la profundidad de los surcos primarios pueden estar relacionadas con la variedad de cacao. Es de tamaño (desde 10 hasta 35 cm de largo) y de forma variable (Figura N° 4). Son de diversos colores al madurar (rojo, amarillo, morado y café); contienen entre 20 y 40 semillas que están cubiertas de una pulpa mucilaginoso de color blanco, cuyos cotiledones pueden ser de color blanco y/o violeta. Las semillas una vez secas alcanzan pesos entre 0,8 y 1,5 g cada una (Braudeau, 1981).

La planta proveniente de semilla presenta un tronco vertical (ortotrópico) que puede desarrollarse en forma muy variada dependiendo de las condiciones ambientales, el cual empieza su etapa de producción a los tres años después de establecido en el campo. Las plantas de origen clonal obtenidas mediante injerto o estacas presentan una conformación diferente sin el predominio de un eje principal (plagiotrópico) y empiezan su producción a los dos años (Adriazola, 2003) (Figura N° 1).



**Figura N° 1:** Fotografía de un Híbrido Bioversity/UNAS  $\pi 16$  de cacao mostrando frutos inmaduros. E.E.A. Tulumayo – Tingo Maria.

## **2.2 Importancia del cacao**

El cacao es un árbol perenne de zonas tropicales húmedas. Actualmente, se ha convertido en un cultivo económicamente muy importante para los pequeños agricultores de las zonas tropicales y subtropicales en el mundo. Asimismo, es un producto básico agrícola o *commodity* negociado en la bolsa de valores y juega un rol muy importante en el sector exportador e industrial de varios países contribuyendo al aumento de sus divisas (Braudeau, 1981; DGPA, 2007; Guiltinan *et al.*, 2008).

Del cacao (principalmente semillas y fruto) se derivan diversos productos (DGPA, 2007):



- Polvo de cacao (cocoa), ingrediente para bebidas, postres de chocolate, helados, *mousse*, tortas, galletas, confitería, etc.
- Pasta o licor de cacao, se utiliza para elaborar el chocolate.
- Manteca de cacao, elaboración de chocolate y confitería, industria cosmética (cremas humectantes y jabones) e industria farmacéutica.
- Pulpa de cacao, producción de bebidas alcohólicas y no alcohólicas
- Cáscara, alimentación de animales.
- Cenizas de la cáscara, puede ser utilizado para la elaboración de jabón y como biofertilizante.
- Jugo de cacao, a partir de la pulpa del fruto, elaboración de vino, vinagre, licores y mermeladas.

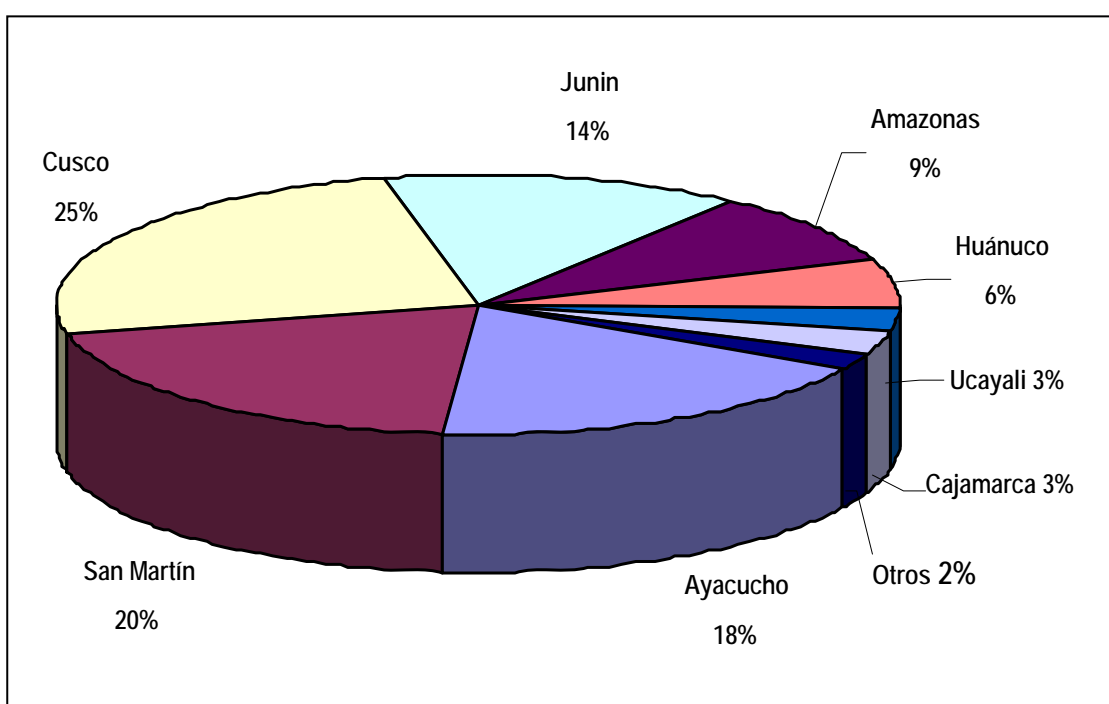
Se cultiva principalmente en África del Oeste, América Central, Sudamérica y sureste de Asia. Según la producción anual los ocho países productores en el mundo son (en orden descendente): Costa de Marfil, Ghana, Indonesia, Nigeria, Camerún, Brasil, Ecuador y Papua Nueva Guinea (ICCO, 2008). Estos países representan el 90% de la producción mundial.

La producción de cacao se produce en su mayor parte en países subdesarrollados debido a sus condiciones edafoclimáticas adecuadas; sin embargo, su consumo se da principalmente en países desarrollados, por ejemplo: los Estados Unidos, Alemania, Francia y el Reino Unido, que abarcan más del 50 % de la demanda mundial (DGPA, 2007). Asimismo, los compradores en los países consumidores son los principales transformadores y productores del chocolate. Dominan la transformación y la producción de chocolate pocas compañías transnacionales.

Actualmente, existen valores agregados al cacao que lo distinguen de los productores comunes. La demanda del cacao orgánico, que incluye un componente “saludable y ecológico”. Tiene un mayor costo que es gustosamente asumido por los consumidores y redundando en los correspondientes productores. Por otro lado, países productores como el Ecuador, están produciendo chocolates con denominación de origen, de alta

calidad y con alto contenido de cacao (60-80%) cuyo precio en el mercado es muy cotizado. En el aspecto social, el sello de certificación de Comercio Justo o “*Fair Trade*”, mejora la posición de los productores desfavorecidos en los países en vías de desarrollo.

El cacao que se produce en el Perú es del tipo aromático y está caracterizado por su alto contenido de grasa (hasta un 57%). Esto le confiere un mayor valor comercial en el mercado internacional (Arca, 2000). Según estadísticas hasta el año 2006 del DGIA-MINAG (DGPA, 2007), la producción del cacao por departamentos se distribuye porcentualmente como se muestra en la Figura N° 2.



**Figura N° 2:** Producción de cacao en el Perú por departamentos, año 2006 (DGPA, 2007).

En términos globales, en todo el Perú se producen alrededor de 29 900 TM de grano de cacao en aproximadamente 54 000 hectáreas. El rendimiento promedio nacional es de aproximadamente 551 kg/ha (DGPA, 2007). Dado que este cultivo se presenta como una actividad promisorio y una buena opción para la agricultura peruana, es necesario incrementar la producción a

través del aumento de áreas, mejora de las técnicas de cultivo y la obtención de variedades con caracteres de interés.

Entonces, se tiene un largo camino por recorrer ya que la productividad nacional en las parcelas de cacao no ha observado un crecimiento significativo en los últimos siete años (IICA, 2006). Esto se ha debido principalmente al manejo de plantaciones con baja tecnología que en su mayoría es conducida por pequeños agricultores que no cuentan con liquidez para la adquisición de insumos o acceder a servicios de asistencia técnica. En consecuencia, la agroindustria nacional importa un promedio de 3 500 TM al año de grano de cacao debido a que la producción nacional no consigue cubrir la demanda. Su destino es la transformación a manteca de cacao y en menor escala a otros productos derivados, con la finalidad de exportarlos (DGPA, 2007).

Se ha promocionado el cultivo cacao como una alternativa al cultivo ilegal de la hoja de coca en muchas regiones pobres del Perú, por ejemplo en la región de Huánuco; sin embargo, este proyecto de sustitución ha conllevado diversas implicancias sociales y los esfuerzos nacionales e internacionales hasta el presente no han logrado éxitos contundentes.

El cultivo del cacao no juega únicamente un rol económico, sino que proporciona beneficios ambientales ya que es considerado uno de los componentes para el sostenimiento ecológico de sistemas agroforestales. Posee efectos restauradores del suelo porque permite la conservación del suelo y el agua, controlando su erosión por medio del acolchado y cobertura de la hojarasca. Además es hábitat temporal de algunas especies migratorias de aves y provee de sombra y refugio a mucha fauna y microfauna (Guiltinan, 2007).

Adicionalmente, al cacao se le atribuyen propiedades terapéuticas debido a su alto contenido de flavonoides o polifenoles, representados en su mayoría por la epicatequina y la catequina. Estos compuestos con actividad antioxidante están en un porcentaje que varía de 12 al 18%. Producirían

efectos positivos en el sistema cardiovascular, reducirían riesgos de cáncer y niveles de estrés. También contiene cerca de un 1,5 a 3% del alcaloide teobromina, el cual es un estimulante del sistema nervioso central, actividad similar a la de la cafeína (Braudeau, 1981). La teobromina es usada como un diurético y como reductor de la presión sanguínea, dado que dilata los vasos sanguíneos (Dias, 2001; Adriazola, 2003; ICCO, 2008).

Entre las cualidades “negativas” que se atribuyen al chocolate, tenemos a las alergias y la obesidad; pero éstas no se deben directamente al cacao, sino a los productos adicionales que se agregan en su elaboración, tales como la leche, nueces, azúcares, etc. (ICCO, 2005). Por ello, muchos consumidores están tendiendo a adquirir los chocolates con alto contenido de cacao, en porcentajes que varían de 60% hasta el 80%.

El cacao no es la única que posee diversos usos. Podemos mencionar a *Theobroma grandiflorum* (“cupuaçu” en portugués) y *Theobroma bicolor* (“patashte” en México), cuyos frutos tienen más pulpa que semilla y son utilizados para la industria alimentaria, principalmente bebidas y confitería, aunque no a gran escala (Braudeau, 1981; Dias, 2001).

## 2.3 Taxonomía y Nomenclatura

La clasificación del cacao según Cronquist, es la siguiente (Arca, 2000; Dias, 2001):

Reino:	Plantae
División:	Magnoliophyta
Clase:	Magnoliopsida
Orden:	Malvales
Familia:	Sterculiaceae
Sección:	Theobroma
Género	Theobroma
Especie:	<i>Theobroma cacao</i> L.

Muchos autores (Motamayor y Lanaud, 2002; Sounigo *et al.*, 2005; Lachenaud *et al.*, 2007; Guiltinan *et al.*, 2008) han reportado al *T. cacao* como perteneciente a la familia Malvaceae *sensu lato* y por tanto ya no dentro de las Sterculiaceae. Según el Royal Botanical Garden of Edinburg, la familia Malvaceae *sensu lato* estaría conformada por las familias Malvaceae *sensu stricto*, Sterculiaceae, Bombacaceae y Tiliaceae. Esto nos lleva a mostrar el Sistema de clasificación APG (del inglés “Angiosperm Phylogeny Group”):

Clado 1: Angiospermas  
Clado 2: Eudicots  
Clado 3: Core Eudicots  
Clado 4: Rosids  
Clado 5: Eurosids II  
Orden: Malvales  
Familia: Malvaceae  
Género: Theobroma  
Especie: *T. cacao* L.

Subdividiendo la familia Sterculiaceae, ésta agrupa 50 géneros, siendo uno de los más importantes el género *Theobroma*. Ésta a su vez comprende 22 especies, que están repartidas en 6 secciones reconocidas por Cuatrecasas, 1964 (citado por Whitlock y Baum, 1999; Dias, 2001; Rondón y Cumana, 2005), mostradas en la Tabla N° 1.

El estudio molecular del gen Vicilin para determinar la filogenética de los géneros *Theobroma* y *Herrania* (Whitlock y Baum, 1999), confirmó a la mayoría de especies por secciones, a excepción de la sección *Andropetalum* que resultó agrupada dentro de la sección *Glossopetalum*.

En la historia de la clasificación intraespecífica de *T. cacao*; Morris, en 1882 (citado por Braudeau, 1981), fue el primer botánico en diferenciarla en “cultivares”, proponiendo dos grandes grupos: Criollo y Forastero. En años subsiguientes fue reafirmada por otros autores (Braudeau, 1981). Cheesman

(1944) sugirió la siguiente clasificación en cuatro grupos básicos: Criollo centroamericano, Criollo sudamericano, Forasteros amazónicos y Trinitarios.

**Tabla N° 1:** Clasificación en secciones taxonómicas del género *Theobroma* y ubicación de la especie de cacao.

Sección	N° especies	Especies
Andropetalum	1	<i>T. mammosum</i>
Glossopetalum	12	<i>T. angustifolium</i> , <i>T. canumanense</i> , <i>T. chocoense</i> , <i>T. cirmolinae</i> , <i>T. grandiflorum</i> , <i>T. hylaeum</i> , <i>T. nemorale</i> ., <i>T. obovatum</i> , <i>T. simiarum</i> , <i>T. sinuosum</i> , <i>T. stipulatum</i> y <i>T. subincanum</i>
Oreanthes	5	<i>T. bernoullii</i> , <i>T. glaucum</i> , <i>T. speciosum</i> , <i>T. sylvestre</i> y <i>T. Velutinum</i>
Rhytidocarpus	1	<i>T. bicolor</i>
Telmatocarpus	2	<i>T. gileri</i> y <i>T. Microcarpum</i>
Theobroma	1	<i>T. cacao</i>

Cuatrecasas (1964) (citado por García, 2000 y Dias, 2001) propuso dividir a *T. cacao*. en dos subespecies: *T. cacao subespecie cacao* para el cacao de América Central y México y las variedades cultivadas derivadas representarían el grupo del cacao Criollo. Mientras que el cacao de Sudamérica fuera clasificado como *T. cacao ssp. sphaerocarpum* y sus cultivares serían representantes del grupo Forastero. Sin embargo, ésta propuesta no tuvo el respaldo necesario y se continuó utilizando la terminología por razas o tipos: Criollo y Forastero (Dias, 2001).

Por otro lado, en la Tabla N° 2, se muestra la clasificación propuesta por Lachenaud (1997) sustentándose sobre la base de estudios moleculares y argumentos paleoclimáticos, paleográficos y geobotánicos. De esta manera, se aportó una clasificación en grupos o compuestos germoplásmicos con su

respectiva distribución geográfica. Excluye aquí al Trinitario, por ser un híbrido entre Criollo y Forastero.

Motamayor *et al.* (2002) en un estudio molecular usando RFLP y SSR, llegaron a la diferenciación del cacao de tipo Criollo en dos: Criollos Ancestrales y Criollos Modernos. Ambos grupos mostraron las características morfológicas para Criollo propuestas por Cheesman (1944). Sin embargo, el muestreo para la primera población involucró individuos de lugares donde el flujo genético entre arboles Criollos y Trinitarios o Forasteros estuviera ausente o limitado por la ausencia de estos dos últimas razas. En cambio, los individuos Criollos Modernos, se colectaron de campos de cultivo recientes con una gran probabilidad de introducción de Trinitarios o Forasteros.

**Tabla N° 2:** Grupos de cacao y distribución geográfica (Lachenaud, 1997 y Motamayor *et al.*, 2002)

<b>Grupo de Cacao</b>	<b>Distribución Geográfica</b>
Criollo ancestrales	Sur de México
Criollos modernos	América central; Colombia y Venezuela
Amazonas o Forastero del Alto Amazonas	Perú, Ecuador, Colombia, Bolivia y Brasil
Guyanas o Forastero del Bajo Amazonas	Meseta de las Guyanas, Venezuela, Surinam, Guyana Francesa y Brasil
Nacional	Zona costera del Ecuador

Se ha confeccionado un mapa de Sudamérica y parte de Centroamérica, que muestra mediante círculos las regiones donde se han encontrado las variedades indicadas (Figura N° 3).

Motamayor *et al.* (2008) analizaron 1241 individuos silvestres y cultivados de cacao de 12 países con marcadores microsatélites; la mayoría de ellos provenientes de Perú, Brasil y Ecuador. Los resultados obtenidos en este trabajo, inspiran la propuesta de una nueva clasificación del germoplasma de

cacao. Ésta describe 10 grandes grupos: Marañon, Curaray, Criollo, Iquitos, Nanay, Contamana, Amelonado, Purús, Nacional y Guyana. Esta nueva agrupación podría reflejar de manera más certera la diversidad genética disponible para los fitomejoradores, y podría ser de mayor utilidad que la tradicional clasificación de Criollo, Forastero y Trinitario.

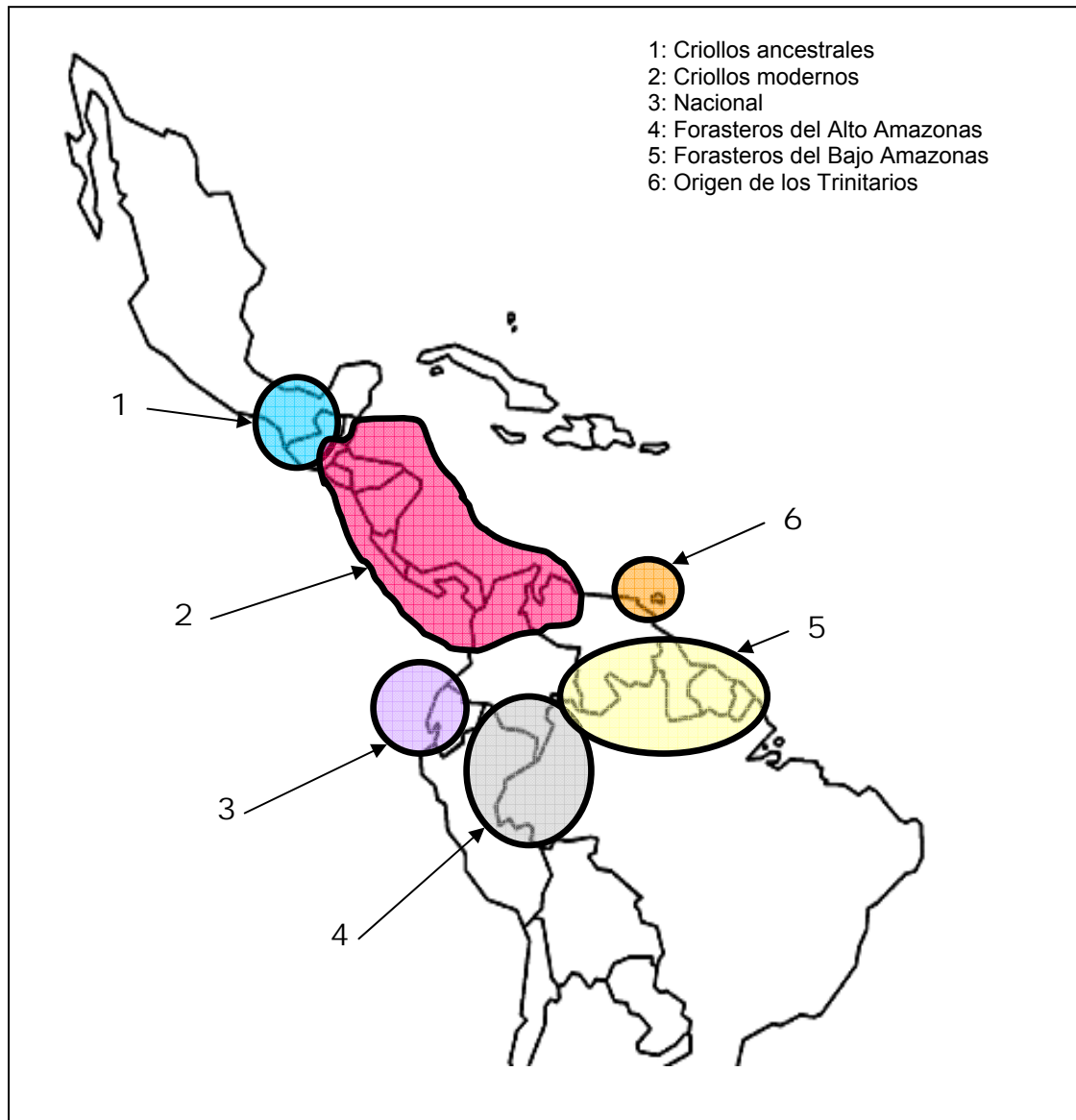
Durante miles de años, el cacao ha sido cultivado en América y todos los tipos o razas que se encuentran hoy no son más que variedades o cultivares de los que apenas es posible conocer su origen y su historia. Se han producido mutaciones, se han hecho selecciones y han ocurrido numerosos cruzamientos naturales (Braudeau, 1981). Entonces, podemos resumir que a pesar de la dificultad de hallar un consenso en la clasificación del cacao, ya que también va de la mano con el tema de su origen y domesticación, se han logrado avances en ese sentido.

## **2.4 Características de las principales razas de cacao**

### **2.4.1 Cacao Criollo**

Es conocido como cacao “fino” de aroma y sabor, produciendo el chocolate de mejor calidad. Se le encuentra en el norte de Sudamérica y Centroamérica. Caracterizado por un fruto con frecuencia alargado, con punta pronunciada, doblada y aguda; la superficie es generalmente rugosa, delgada, de color verde frecuentemente con salpicaduras de rojo a púrpura oscuro y marcada por 10 surcos profundos; los granos son grandes, gruesos, de sección casi redonda con los cotiledones blancos o de un color rosado. De este tipo se obtiene el chocolate reconocido de más alta calidad; sin embargo, tiene baja producción y es muy susceptible a enfermedades. Se cultiva en México, Venezuela, Colombia, Nicaragua, Guatemala, Trinidad, Jamaica y Granada, en la zona del océano Índico y en Indonesia. Representa, como mucho, el 10% de la producción mundial (Wood, 1975; Adriazola, 2003).





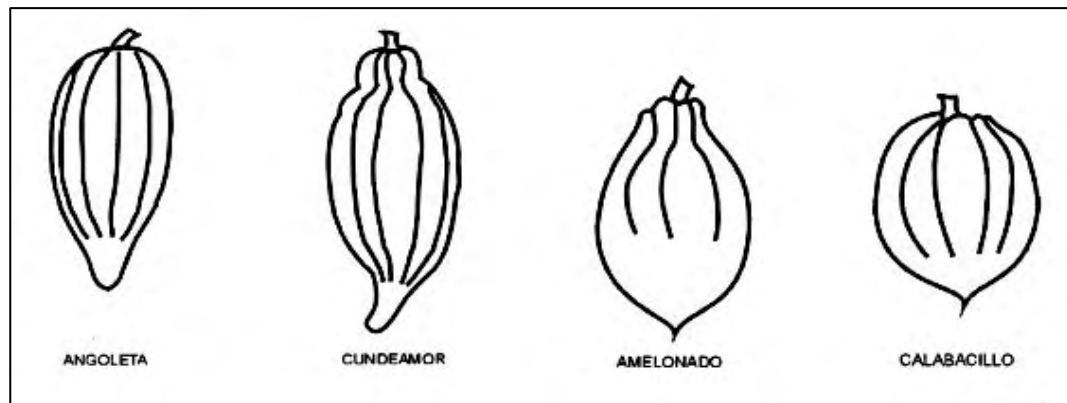
**Figura N° 3:** Mapa de Latinoamérica mostrando la distribución postulada para los diversos grupos de cacao (a partir de Lachenaud, 1997 y Motamayor *et al.*, 2002).

#### 2.4.2 Cacao Forastero

Proveniente de la Cuenca Amazónica, el fruto es generalmente de forma ovalada y corta, de color verde o amarilla cuando madura, con una superficie lisa. Pericarpio espeso y difícil de cortar a causa de un mesocarpo fuertemente lignificado; granos pequeños y más o menos aplastados y tienen un color entre púrpura claro y oscuro. Posee un grado elevado de taninos, que le transfiere el sabor amargo al chocolate. Es por esto que se le

considera un cacao “corriente”. Forma un complejo grupo morfológico tanto en sus formas silvestres como cultivadas. Dada su alta producción, el tipo forastero domina la producción mundial. Se cultiva en África y Sudamérica.

Van Hall (1932) citado por García (2000); propuso una subdivisión del grupo Forastero, tomando como referente principal la forma del fruto (Figura N° 4), del cual resultó en cuatro categorías: Angoleta, Cundeamor, Amelonado y Calabacillo. Sin embargo, Cheesman (1944) y posteriormente Allen y Lass (1983) citados por García (2000); han sugerido que esta terminología deberían ser desechadas, ya que han generado reportes equivocados o a veces contradictorios, dada la alta variabilidad que se presenta.



**Figura N° 4:** Esquema de las cuatro formas principales de fruto del cacao Forastero (García, 2000)

### **2.4.3 Cacao Trinitario**

Se cultiva en los todos países donde anteriormente se cultivaron los criollos, México, Trinidad, Colombia, Venezuela. Asimismo, se están introduciendo en los demás países productores. Se trata de una población híbrida que se originó de una cruce entre un Forastero (de la variedad amelonado) de las Guyanas y Criollo de Venezuela (García, 2000). Posee una alta heterogeneidad variable debido a su origen híbrido. Fue anteriormente clasificado como un tipo de Forastero, es de origen reciente y puede ser reproducido artificialmente. Producen un cacao de calidad aunque inferior al Criollo (Wood, 1975; Adriazola, 2003).

#### **2.4.4 Cacao “Nacional”**

Oriundo de Ecuador y es responsable de la reputación de la calidad del cacao de este país, considerado como un cacao “fino”, con intenso aroma a chocolate y con rasgos de olores a flores y especias. A pesar de estas cualidades organolépticas, no es de origen Criollo sino Forastero Amazónico (Braudeau, 1981; Adriazola, 2003). Su calidad es denominada “Arriba”. Está siendo desplazado por los Trinitarios. Las mazorcas son de color verde, grandes, ovales, de forma amelonada con una ligera estrangulación en la base, de punta embotada. El pericarpio es espeso y marcado con surcos profundos; su superficie es bastante rugosa. Las semillas son grandes, de color violeta mas claro que la de los demás Forasteros. En alguna literatura lo incluyen en un grupo de cacao llamado “Refractario”.

#### **2.5 Centro de Origen.**

El género *Theobroma* se encuentra en estado natural en los pisos inferiores de las selvas húmedas de América tropical. Se extiende desde la llanura de la cuenca y tributarios del Amazonas hasta el sureste de México. Prospera entre los 18° N (otros autores señalan desde los 15° N) hasta los 15° S del Ecuador a una altitud inferior a 1,250 m.s.n.m. Excepcionalmente se encuentra en latitudes subtropicales a 23° y 25° S; por ejemplo, en esas latitudes se ubica el Estado de Sao Paulo, Brasil (Adriazola, 2003).

Como cultivo, el cacao se llevó de Brasil a la colonia portuguesa de Príncipe en 1822 y de ahí a la isla Sao Tomé en 1830, ambas en el Golfo de Guinea. Posteriormente, gracias a la corona británica, Ghana obtiene el cacao en 1869 y por el año de 1951 se extiende al oeste de África, donde actualmente la producción mundial alcanza un 60 %.

El centro de origen de *T. cacao* se ha considerado un tema controversial (Dias, 2001; Motamayor *et al.*, 2002) debido a las siguientes razones:

- Importancia de las variedades domesticadas en Centroamérica.
- Dispersión de la especie por el hombre o de manera natural.
- Mayor diversidad genética en especies de la zona alta de la cuenca del río Amazonas (involucra a Brasil, Perú, Bolivia, Ecuador, Venezuela y Colombia).

En la cuenca alta del río Amazonas, es en donde se presenta la mayor variabilidad de esta especie. A *grosso modo* y como se describe en la Tabla N° 2, las poblaciones de Centroamérica y norte de Sudamérica conforman el grupo racial llamado “Criollo” y aquellos de la cuenca del Amazonas forman el grupo “Forastero”.

Se han planteado tres hipótesis (Dias, 2001; Motamayor y Lanaud, 2002), mutuamente excluyentes y conflictivas, que proponen el origen y dispersión de estas poblaciones:

- Hipótesis de dispersión Sur a Norte: supone que el cacao, cuyo centro de origen sería la cuenca alta del Amazonas, habría sido llevado por humanos a través de los Andes hacia Centroamérica, siendo aquí donde se domesticó.
- Hipótesis de dispersión Norte a Sur: respalda un origen Centroamericano y dispersión del cacao en dirección opuesta a la anterior hipótesis; el centro de origen habría sido Centroamérica y a partir de aquí los cacaos habrían sido transportados por Amerindios hacia Sudamérica.
- Desarrollo simultáneo e independiente de dos poblaciones de cacao, separados por el Istmo de Panamá, y que derivó en cacao Criollo en Centroamérica y Forastero en Sudamérica.

Las evidencias antropológicas, históricas, paleontológicas y biogeográficas, junto con los estudios moleculares de genética de poblaciones han reforzado la hipótesis de dispersión Sur a Norte (Dias, 2001; Motamayor *et al.*, 2002).

Dias (2001) reforzaría la primera hipótesis: postulando que en épocas precolombinas se transportaron semillas de cacao Criollo y Forastero por humanos desde el centro de origen del cacao (cuencas del Alto Amazonas y del Orinoco), a lo largo de los valles de los Andes hasta Centroamérica. En el imperio Maya, se inició un proceso consciente de selección dirigida a obtener alta calidad de chocolate, resultando en una variedad domesticada perteneciente a los Criollos, desplazando al cacao Forastero.

Esta discusión sobre el origen y domesticación del cacao amplía el entendimiento del proceso evolutivo de la especie, con significativos impactos en la conservación y el mejoramiento genético.

## **2.6 Recursos genéticos del cacao.**

La diversidad biológica comprende el conjunto de poblaciones nativas y domesticadas, de distinto origen genético y grado evolutivo, que ocupan nichos ecológicos específicos y, que teniendo características semejantes y distintivos, se muestran variables (García, 2000).

Entonces, los recursos genéticos son una fracción de esta diversidad biológica, que tienen un importante valor actual o potencial para el establecimiento de la agricultura que hoy sustenta la alimentación de la población mundial. Debido a su gran importancia, es que el hombre debe conocerlos, manejarlos, mantenerlos y utilizarlos racionalmente (Jaramillo y Baena, 2000).

Es necesario enfatizar que los recursos genéticos no son renovables, por ello, su conservación es fundamental (Dias, 2001). La pérdida de éstos se denomina erosión genética y se debe principalmente a acciones antrópicas. La deforestación, expansión del área agrícola y el mejoramiento y selección de nuevas variedades contribuyen a la disminución del acervo genético presente en el germoplasma.

La conservación de germoplasma es una labor continua e implica inversiones importantes en tiempo, personal, instalaciones y operación, justificables en función de las necesidades. Las razones y el material que se van a conservar deben definirse con criterios lógicos, científicos y económicos como la necesidad, el valor y uso de las especies, la factibilidad, el costo y beneficio (Jaramillo y Baena 2000).

Esta fuente potencial de variabilidad genética es de gran utilidad para los programas de mejoramiento de cacao. También, desempeñan el rol fundamental para la evolución natural de la especie. Pueden proporcionar la materia prima necesaria para crear nuevos cultivares, los cuales pueden ser más productivos, más adaptados a la región y clima, y más resistentes a enfermedades y plagas más importantes (Dias, 2001). Es obvio que estos recursos orientados hacia el cacao seguirán siendo de enorme importancia en la sostenibilidad del cultivo en el Perú (IICA, 2006).

Sin embargo, dada su ecología y biología, la incorporación de tales recursos genéticos de cacao en los programas de mejoramiento es un asunto prolongado y complicado. De hecho, las poblaciones de cacao silvestre se encuentran en zonas difíciles de acceder (selvas tropicales de Sudamérica), limitando las exploraciones y el cuidado de los materiales colectados. En este dilema, es necesario una planificación y administración apropiadas del material genético ya presente en los bancos de germoplasma y determinar de forma exhaustiva la diversidad y sus rasgos importantes mediante su caracterización y evaluación. Esto incrementará su relevancia y utilidad (Engels y Visser, 2007).

Debido a que la semilla del cacao es de tipo recalcitrante, su conservación se realiza en campos *ex - situ* o *in - situ* de propagación clonal (bancos de germoplasma). Diversos países productores de cacao conservan de esta manera su germoplasma: Brasil, Costa de Marfil, Costa Rica, Ecuador, Ghana, Guyana Francesa, Malasia, Perú, Trinidad y Tobago, Venezuela, etc. (García, 2000; Martínez, 2007).

Los bancos pueden ser de dos tipos (García, 2000): Colecciones Base, que albergan un gran número y diversidad de accesiones o Colecciones Activas, con menor número y diversidad de accesiones. Bioversity International (antes IPGRI) considera a las siguientes colecciones base como “colecciones internacionales o universales” :

- The International Cocoa Genebank at Trinidad (ICG,T), manejado por el CRU en Trinidad y Tobago.
- The International Cacao Collection at CATIE (IC3), manejado por el Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Costa Rica.

La Universidad Nacional Agraria de la Selva (UNAS) de Tingo María (Huánuco), tiene a su cargo un Banco de Germoplasma que es una colección activa de cacao. Éste le fue transferido desde el año 1990 por Programa de Naciones Unidas para el Desarrollo (PNUD) con la finalidad de conocer, evaluar y generar semilla élite. El pool genético ahí está conformado por aproximadamente 160 accesiones: colección Internacional o Introducida, colección Huallaga y colección Ucayali – Urubamba (García, 2000; IICA, 2006).

## **2.7 Mejoramiento Genético**

El cacao es de naturaleza diploide y posee  $2n=2x=20$  cromosomas (Dias, 2001; Motamayor y Lanaud, 2002). Es una especie principalmente alógama, monoica con flores hermafroditas y dependiendo del genotipo posee un sistema de autoincompatibilidad. Ésta característica permite que el cacao asegure un alto grado de diversidad genética en sus poblaciones segregantes.

La larga duración del ciclo de vida del árbol de cacao prolonga considerablemente los programas de mejoramiento: toma al menos 15 años desde la siembra de los ensayos varietales hasta su validación. También, es

necesario añadir el tiempo impuesto por cuarentena obligatoria cuando existe el traslado de material vegetal entre los continentes con el fin de evitar la diseminación de enfermedades (Adriazola, 2003).

Todo esto explicaría porque usualmente se requieren de varias décadas desde la colección de recursos genéticos y su uso práctico por los agricultores, siempre y cuando tales recursos no se pierdan durante los años iniciales, durante las etapas de mantenimiento, caracterización y evaluación.

El mejoramiento genético demanda una mayor diversidad genética en las colecciones, siendo la caracterización y evaluación de éstas, una de las mayores prioridades. En forma similar que otros cultivos perennes, el cacao depende de la disponibilidad de material genético de origen silvestre para la obtención de variedades importantes (especialmente para aspectos de resistencia a enfermedades).

Sin embargo, la mayor parte del germoplasma está insuficientemente caracterizado. En los casos más favorables sólo conocemos datos ecogeográficos (datos de pasaporte) y algunos caracteres fenotípicos. Las mejoras en la caracterización genética de los materiales y la masiva disponibilidad de marcadores de calidad contribuirán a un uso más racional de los recursos genéticos. Los investigadores y los mejoradores estarían interesados en caracteres específicos y en el conocimiento detallado de tales caracteres, incluyendo la información de tipo molecular y bioquímico, mientras que los agricultores están más interesados en propiedades agronómicas (Engels, 1986; Engels y Visser, 2007).

Comúnmente, se han usado dos sistemas para seleccionar clones de cacao: el primero es la selección individual de árboles, que luego son multiplicados vegetativamente, y el segundo, la obtención de plantas por polinización abierta o controlada (Adriazola, 2003).

Mediante la utilización de marcadores moleculares, se puede acortar el tiempo de obtención de nuevas variedades de cacao. Esto se logra a través



de la selección asistida por marcadores o *Marker Assisted Selection* (MAS), haciendo más eficiente la selección de plantas y de variedades con las características de interés. Son técnicas que puede realizarse a nivel de plántulas de vivero o previvero, lo cual ayuda a reducir significativamente el tiempo y los costos de manejo de un programa de mejoramiento (Guiltinan, 2007).

Dias (2001) menciona que las características principales a seleccionar en un programa de mejoramiento genético de cacao son las siguientes: mayor peso seco de semilla, resistencia a enfermedades locales, disminución del costo de producción (por ejemplo: índice de mazorca), aumento de la calidad del producto final (mayor peso de grano, menor porcentaje de testa, gran contenido de grasa).

## **2.8 Caracterización de los Recursos Genéticos**

Se refiere al uso de procesos y métodos para generar información sobre el pool genético mantenido en el banco. Es la descripción de la variación que existe en una colección de germoplasma y permite diferenciar a las accesiones que la integran (Arciniegas, 2005). La evaluación y caracterización puede ser llevada a cabo usando los marcadores morfológicos, marcadores bioquímicos (apartado 2.9) y/o marcadores moleculares (apartado 2.10).

La caracterización sirve para múltiples usos tales como: proporcionar un mejor conocimiento del germoplasma, identificar duplicados, identificar genotipos faltantes en las colecciones, facilitar la planificación de nuevas colectas e introducciones, permitir el establecimiento de colecciones núcleos. En la identificación de especies, familias y géneros de plantas los caracteres morfológicos han sido muy usados, constituyéndose éstos en una herramienta útil e indispensable para realizar numerosos estudios en genética de poblaciones y agricultura (Arciniegas, 2005).

La caracterización morfológica se basa en atributos o estados observables o medibles referentes a la forma, aspecto, estructura o comportamiento del todo o partes del ciclo vital del individuo. Pueden consistir en el empleo de características botánicas, fenológicas y agronómicas. Se conoce que tienen variación entre individuos y frecuentemente son convertidos en descriptores estandarizados (Saad y Rao, 2001). Se eligen aquellos caracteres de alta heredabilidad; sin embargo, estos rasgos resultan de la expresión de un número limitado de genes y son dependientes de la interacción con el medio ambiente (Dias, 2001; Arciniegas, 2005).

Dentro de los caracteres morfológicos que definen las características del género *Theobroma* existen diversas instituciones que han propuesto una lista de descriptores morfológicos para la identificación y evaluación del germoplasma de cacao. Por ejemplo, el *International Board for Plant Genetic Resources* (IBPGR) ha seleccionado 65 descriptores, en tanto que otros autores propusieron una lista corta de 26 descriptores morfológicos y el *Centre de coopération Internationale en Recherche Agronomique pour le Développement* (CIRAD) emplea 24 descriptores para la caracterización.

En el CATIE se han llevado diversas caracterizaciones morfológicas de su material germoplásmico (Engels, 1986; Arciniegas, 2005; Martinez, 2007). En Tingo Maria - Perú, mediante trabajos de tesis se han caracterizado las accesiones del banco de germoplasma. Se evaluaron características morfológicas o botánicas, biométricas, compatibilidad, calidad del grano, reacción frente a enfermedades, etc. (Bartra, 1993; López, 1993; Piñan, 1993; Gonzales, 1996; Vidal, 1997; García, 1998; Panduro, 1998; Rengifo, 2002; Adiazola, 2003; Guerrero, 2005; Aycachi, 2008).

El conocimiento y evaluación de los rasgos morfológicos son muy importantes y no pueden ser reemplazados por ninguna técnica molecular. Los resultados de los estudios moleculares y bioquímicos deben ser considerados como complementarios a la caracterización morfológica (Karp *et al.*, 1997; Spooner *et al.*, 2005).

En los Anexos N° 1 y 2, se presentan para el cacao una selección de datos morfológicos muy estables y caracteres de interés agronómico, respectivamente (algunos de los más importantes elegidos de la literatura revisada, que serán utilizados para discutir los resultados moleculares).

## **2.9 Caracterización con Marcadores Bioquímicos**

Se basa principalmente en la detección de diferencias de migración entre enzimas cuya función es idéntica, pero producidas por alelos diferentes de un mismo gen. Así, son conocidas también como aloenzimas o simplemente alozimas. Éstas catalizan la misma reacción bioquímica, pero difieren en composición por uno o más aminoácidos, por lo que tienen diferentes carga eléctrica y conformación estructural (Spooner *et al.*, 2005). Debido a estos factores, la migración electroforética de estas moléculas es afectada, y se pueden detectar las variaciones alélicas. Manipulando al substrato y cofactores se pueden provocar reacciones enzimáticas de modo que se produzca coloración, formando bandas visibles donde las alozimas en particular se localizan en el gel de electroforesis. Por lo común, éste es de almidón o poliacrilamida.

Son de característica codominante para la cuantificación de la heterocigosidad, diversidad genética, variación intra e interpoblacional, identificación de clones, análisis de paternidad, etc. Las desventajas de marcadores alozímicos es su escaso número y se encuentran sujetos a modificaciones postranscripcionales. Además, es necesario hacer uso de tejido fresco o congelado apropiadamente para evitar su desnaturalización y la pérdida de la actividad enzimática.

En el cacao se han realizado importantes trabajos de caracterización y estudios de diversidad genética con aloenzimas (Parra *et al.*, 2001; Parra *et al.*, 2002; Sounigo *et al.*, 2005), obteniendo sólidos resultados en caracterización y análisis de diversidad genética.

## 2.10 Caracterización Molecular

Gracias al vertiginoso progreso de la biología molecular, se han desarrollado una serie de técnicas de marcadores moleculares que permiten el análisis de las variaciones genéticas en los organismos. Entre los principales tipos de marcadores moleculares podemos mencionar los siguientes (Karp *et al.*, 1997; Spooner *et al.*, 2005):

- Hibridización de ADN: RFLP y VNTR.
- Basados en PCR, dentro de esta categoría están:
  - Marcadores con iniciadores arbitrarios o semiarbitrarios, RAPD, DAF, ISSR, AFLP, etc.
  - Marcadores con iniciadores específicos, STS, SSR, SCAR, CAPS, etc.
  - Sitios específicos y detección por desnaturalización del ADN: SSCP, DGGE, TGGE.
  - Sitios específicos y secuenciamiento del ADN: SNP, EST, etc.

Estas técnicas pueden diferir en aspectos tales como: capacidad de amplitud genómica, nivel de polimorfismo detectado, especificidad de locus, reproducibilidad, requerimientos técnicos, grado de inversión financiera y laboriosidad.

La elección del tipo de marcador molecular a usarse puede depender de la disponibilidad de tiempo y fondos para la investigación, limitaciones en el equipamiento, nivel de polimorfismo en la población a estudiar, etc. No se puede decir que existe un único enfoque molecular que permita solucionar los diversos problemas de caracterización y administración en los bancos de germoplasma. Muchas de estas técnicas pueden complementarse una con la otra (Spooner *et al.*, 2005).

Los marcadores moleculares presentan algunas ventajas a diferencia de los morfológicos, y se describen a continuación (Karp *et al.*, 1997; Saad y Rao, 2001; Spooner *et al.*, 2005):

- Son específicos para cada individuo, grupo de individuos, especies o aún de grupos sistemáticos mayores, los convierte en herramientas útiles para el análisis tanto de individuos como de poblaciones.
- Independiente de la influencia del medio ambiente y los efectos pleiotrópicos.
- No están bajo control poligénico, y sujetos a control epistático y modificaciones (tienen plasticidad casi nula).
- Se disponen de un número potencialmente ilimitado de marcadores, proporcionando una gran cantidad de información útil al fitomejorador.
- Los caracteres de ADN pueden ser más fácilmente anotados como estados discretos de alelos o pares de bases de ADN, mientras que los datos morfológicos, bioquímicos y de evaluación de campo son frecuentemente anotados como caracteres continuos y muy variables y no son tan factibles de utilizar en robustos métodos analíticos.
- Muchos marcadores moleculares son neutrales selectivamente, es decir, no están sujetos a una presión de selección.

En el cacao se han logrado utilizar los marcadores RFLP, RAPD, AFLP y SSR para mapeos genético y molecular, mapeo de QTL, diversidad genética, estudio de poblaciones, estudios de flujo génico, estudios filogenéticos, búsqueda de genes de interés, verificar o reconstruir los datos de pasaporte, tasas de mutación en cultivo *in vitro*, etc. (Dias, 2001; Guiltinan, 2007; Zhang *et al.*, 2006; Guiltinan *et al.*, 2008).

Diversos autores (Zhang *et al.*, 2006; Guiltinan, 2007) fundamentan que los marcadores SSR son los más aceptados y confiables para la caracterización molecular de cacao. Esto se debe a que son simples de realizar, posible de automatizar y altamente reproducibles. En recientes convenciones internacionales (Zhang *et al.*, 2006) muchos laboratorios en el mundo llegaron a un acuerdo de usar de 15 iniciadores SSR en la caracterización de

germoplasma de cacao, lo cual permitirá el intercambio de información y la homogenización de las bases de datos. Sin embargo, la utilización de microsatélites tiene sus limitaciones en países en vías de desarrollo, ya que se requiere de un secuenciador de capilares para la detección de las bandas amplificadas.

No cabe duda que la combinación de información procedente de diferentes tipos de marcadores (obtenidos de secuencias no codificantes o de genes funcionales) proporciona la mejor aproximación posible al conocimiento sobre la variación genética (De Vicente *et al.*, 2004). Asimismo, la información molecular sería de poca ayuda si no se cuenta con datos morfológicos y agronómicos. En este contexto, se puede lograr una caracterización preliminar con marcadores moleculares más accesibles, simples de utilizar y de bajo costo pero con alta reproducibilidad, como los ISSR.

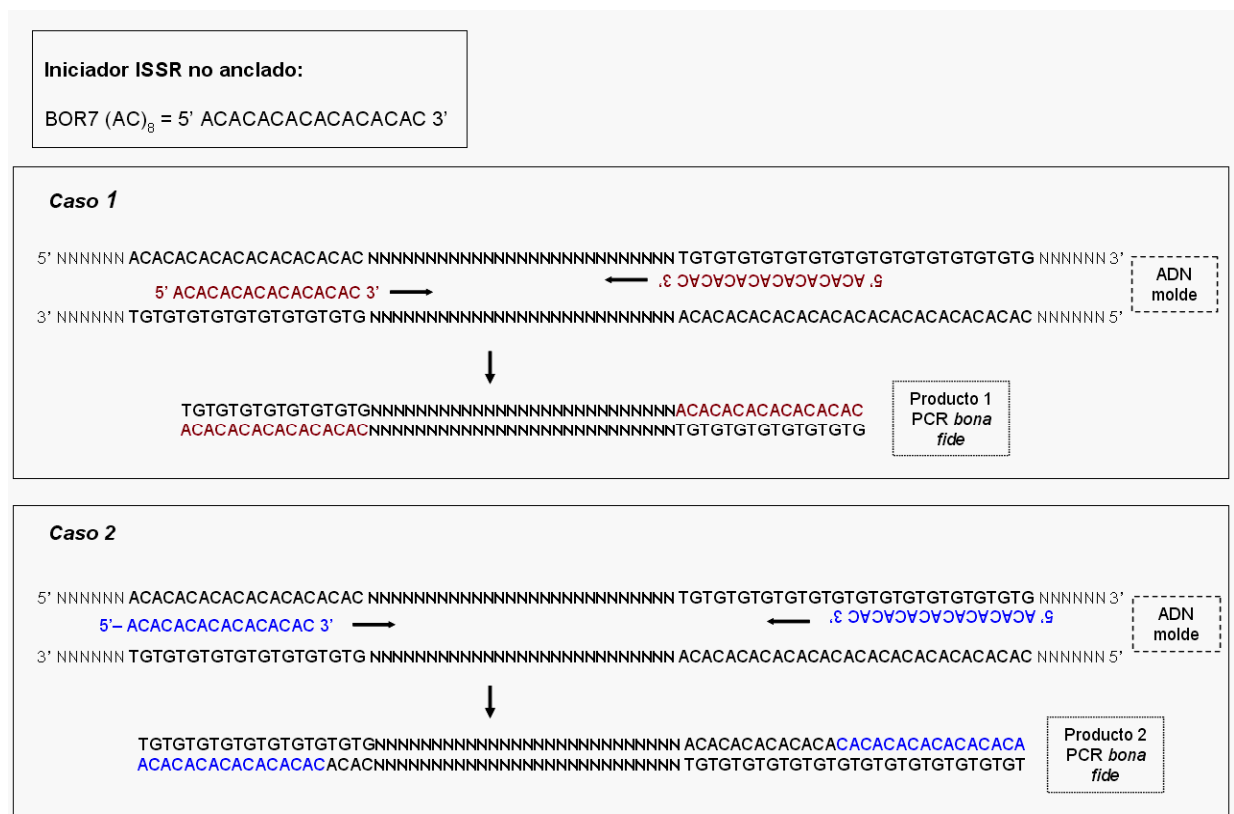
## **2.11 Marcadores moleculares ISSR**

Las siglas ISSR provienen de la frase *Interspread single sequence repeats*, traducido como secuencias intercaladas entre los microsatélites. Fueron desarrollados por Zietkewicz *et al.*, 1994 (citado por Charters y Wilkinson, 2000 y Spooner *et al.*, 2005). Técnica basada en PCR, es de característica semiarbitraria debido a que se utiliza un solo primer de 14 pb o más, complementario a dos microsatélites cercanos presentes en el genoma. La amplificación ocurrirá si ocurre un adecuado alineamiento de los iniciadores y la distancia entre éstos sea de 100 a 2500 pb, o más dependiendo de las condiciones PCR. Por otro lado, la región no será amplificada (banda ausente) si existe divergencia entre los sitios de unión y los iniciadores, si uno de los sitios de unión no existe y/o si existen rearrreglos estructurales (genómicos o cromosómicos).

La hibridación del *primer* está condicionada a su diseño y a la estrictez de la temperatura de alineamiento usada. Son técnicamente simples y no implican

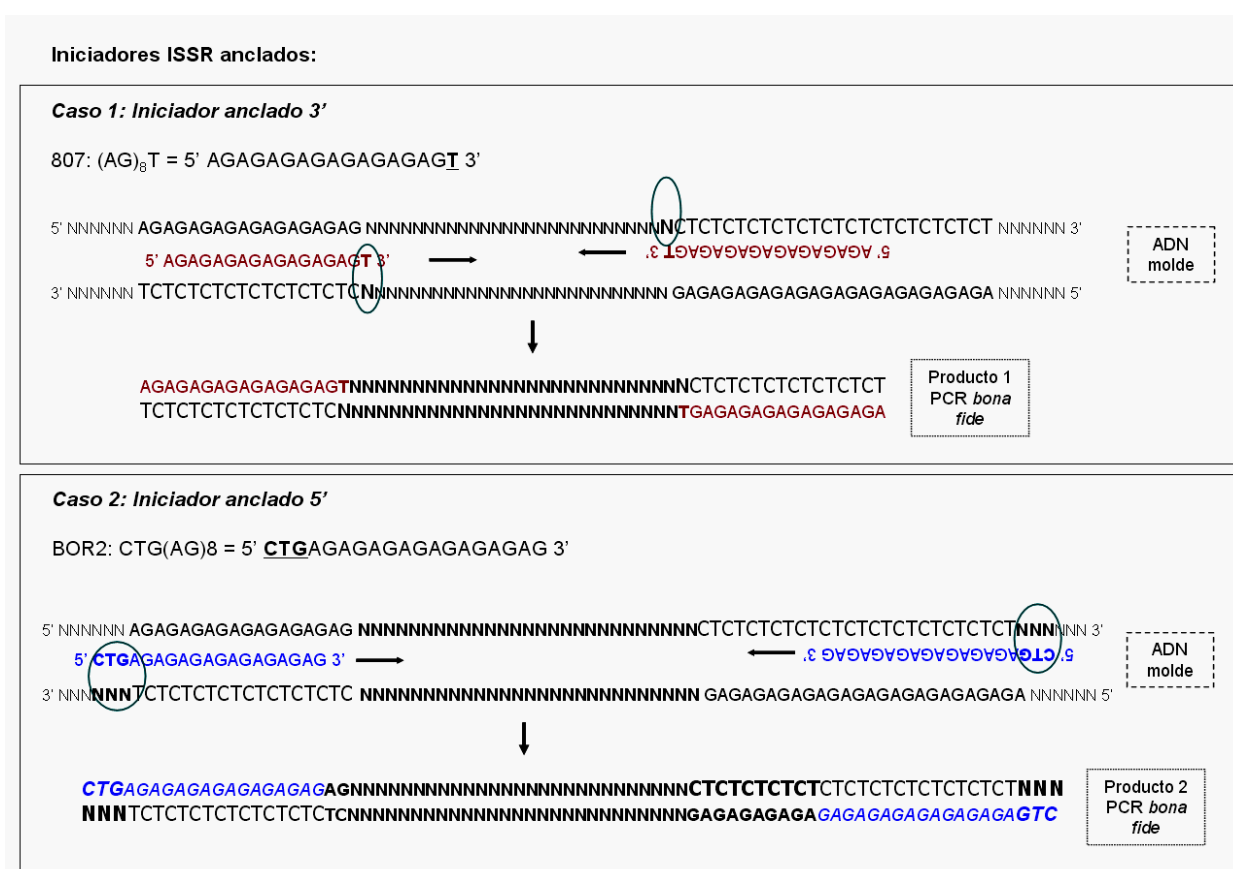
el mayor costo y laboriosidad de los AFLP, pero proporcionan un alto polimorfismo. Debido a que la secuencia del primer es pre-diseñada no requiere el conocimiento previo de las secuencias en el genoma. Tienen una reproducibilidad aceptable y mayor que los RAPD.

Tenemos dos clases de iniciadores: no anclados y anclados. Los primeros poseen secuencia complementaria al microsatélite pero no posee nucleótidos diferentes a éste (Figura N° 5). Los segundos poseen uno o varios nucleótidos diferentes (en posición 3' o 5') que posibilitan el alineamiento del iniciador en una posición específica con respecto al microsatélite (Figura N° 6). Charters y Wilkinson (2000) usaron el ISSR con iniciadores anclados para demostrar su utilidad en la diferenciación de accesiones y la caracterización genética de germoplasma de cacao.



**Figura N° 5:** Esquema mostrando el alineamiento de los iniciadores ISSR no anclados con respecto a un ADN molde (elaboración propia). Los *primers* ISSR anclados se muestran en color rojo y azul. Los casos se explican en el texto.

En la Figura N° 5, tanto en el caso 1 como en el 2 se presentan dos lugares diferentes de alineamiento para un mismo *primer* no anclado; es decir, que los iniciadores, dada su complementariedad total con la secuencia microsatélite, se podrían alinear a lo largo de éste, pudiendo variar de esta manera la longitud del producto amplificado. Esto está muy influenciado por las condiciones de temperatura que se le proporcione a la reacción. Los ISSR no anclados se han utilizado para medir la relaciones intraespecíficas e interespecíficas, como ha sido demostrado con diferentes especies por Bornet y Branchard (2001).



**Figura N° 6:** Esquema mostrando el alineamiento de iniciadores ISSR anclados con respecto a un ADN molde (tomado y modificado de: De Vicente y Fulton, 2003). Los *primers* ISSR anclados se muestran en color rojo y azul. Los casos se explican en el texto.

En la Figura N° 6 se presentan dos casos: en el caso 1 se trata de un *primer* anclado con un nucleótido T (en posición 3') adicional a la secuencia del



microsatélite (AG). Esto permitiría anclar al iniciador en la región más cercana de la región entre los dos microsatélites involucrados, ya que para que ocurra la hibridación en ambas posiciones, debe haber complementariedad entre el nucleótido N y la Timina 3' del *primer* (señalados con un círculo). El caso 2 ocurre cuando el *primer* anclado posee la secuencia adicional en el extremo 5', por lo cual se obligará al alineamiento en los extremos más alejados de los microsatélites invertidos, generando un producto más extenso que en el generado en el caso 1. Asimismo, en el caso 2, se señala con un círculo un grupo de tres nucleótidos (CTG) del extremo 5' del cebador que deben ser complementarios a los tres nucleótidos (NNN) para que los *primers* se logren hibridarse consistentemente al ADN molde.

Los ISSR son interpretados como marcadores dominantes, es decir, que la presencia de la banda amplificada incluye a ambos de los siguientes caracteres: homocigoto dominante y heterocigoto, que son evaluados como uno solo y la banda ausente es considerada como el carácter homocigoto recesivo. Asimismo, en las bandas individuales pueden estar comprometidos varios productos de amplificación comigrantes (Karp *et.al.*, 1997). En otras palabras, sólo un alelo es detectado, que es el segmento amplificado, mientras que las demás variaciones alélicas son clasificadas conjuntamente como un alelo nulo. Esto trae la desventaja de observarse un bajo contenido de información genética por locus y por consiguiente no es posible determinar algunos parámetros genéticos (Ferreira y Grattapaglia, 1998). En ausencia de un análisis de pedigrí, la identidad de las bandas individuales no se conoce (marcadores anónimos). También, la presencia de una banda de peso molecular aparentemente idéntico en diferentes individuos no es evidencia de que los 2 individuos compartan el mismo fragmento homólogo, lo cual podría ocurrir en individuos lejanamente relacionados. Esta característica es conocida como homoplasia, en el sentido de que estos caracteres son comunes entre los individuos pero no provienen de un origen o ancestro común.

La reproducibilidad de una técnica es clave para su éxito y es por esto que el ISSR, a pesar de su característica de semiarbitrariedad, posee una mayor reproducibilidad que el RAPD. Este último, ha sido siempre cuestionado por la inconsistencia de resultados entre laboratorios que comparten los mismos objetivos. La reproducibilidad de los ISSR no está fuertemente influenciada por los cambios del número de ciclos de amplificación PCR o por las amplificaciones llevadas a cabo en diferentes termocicladores, como sí ocurriría con la técnica RAPD. Como resumen en la Tabla N° 3 se muestran las ventajas y desventajas de los ISSR:

**Tabla N° 3:** Ventajas y desventajas de los marcadores ISSR.

<b>Ventajas</b>	<b>Desventajas</b>
No requieren información previa sobre secuencias.	Marcadores dominantes.
Se puede encontrar variación dentro de regiones únicas del genoma en varios loci simultáneamente.	Puede ser necesaria la electroforesis en gel de poliacrilamida y la detección con tinción de plata.
Tienden a identificar niveles significativos de variación.	No es posible asignar bandas ausentes a un mismo tipo de homología genética.
Específicos de secuencias de microsatélites.	No es posible hacer análisis genético de algunos parámetros de diversidad genética.
Muy útiles para realizar perfiles de ADN, especialmente de especies estrechamente relacionadas.	

## 2.12 Taxonomía numérica

La taxonomía numérica o fenética es un sistema de clasificación en la sistemática biológica. El concepto fue propuesto y desarrollado por Sneath y Sokal (1973). Aunque considerado como método objetivo de clasificación, en la práctica la elección y el peso de las características (morfológicas,

moleculares, etc.) es a menudo guiada por los métodos disponibles e intereses en la investigación. Además, dada que su finalidad es la clasificación de los organismos basada en la similaridad global, frecuentemente no toma en cuenta sus relaciones evolutivas o filogenéticas. Asimismo, los algoritmos fenéticos a menudo no logran reconstruir apropiadamente la historia evolutiva de los organismos estudiados.

Actualmente, el consenso general es que la clasificación taxonómica debería reflejar procesos evolutivos (filogenéticos). En consecuencia, la fenética ha sido grandemente superada por la cladística para la investigación de las relaciones entre las especies. Sin embargo, ciertos métodos fenéticos, tales como *Neighbor Joining*, han sido adoptados por la cladística, como una razonable aproximación de la filogenia, cuando métodos más avanzados (tales como la Inferencia Bayesiana) son muy exigentes en recursos computacionales.

Las técnicas fenéticas incluyen varias formas de agrupamiento y ordenación. Éstas son vías sofisticadas de reducir la variación mostrada entre los organismos a un nivel manejable. En la práctica esto significa la medición de docenas de variables (si bien pueden llegar a más), y la presentación de ellas en gráficos más simples de dos o tres dimensiones. El mayor desafío técnico en la fenética constituye el problema de balancear la pérdida de información en tal reducción de variables a favor de la facilitación de interpretar los gráficos resultantes (Crisci y López, 1983).

### **2.13 Estimación de la Semejanza Taxonómica mediante análisis multivariado.**

Los fragmentos de ADN amplificados por técnicas de marcadores moleculares se visualizan en un gel de electroforesis como bandas de diferentes pesos moleculares. La presencia o ausencia de éstas pueden ser compartidas entre los genotipos (de aquí en adelante llamados *taxa* u OTUs, del inglés *Operational Taxonomic Units*) y por lo tanto se consideran caracteres fenotípicos distintos e independiente de los demás. Son muy

utilizados los valores (1) y (0) del método binario para representar el valor de presencia o ausencia, respectivamente.

Con el fin de determinar el grado de similaridad entre los individuos en estudio y agruparlos en “*clusters*” o ramas comunes (Sneath y Sokal, 1973), se aplican diversos métodos estadísticos multivariados tratándose del análisis de una gran cantidad de variables (loci). La información obtenida permite calcular las distancias genéticas y generar representaciones gráficas. Asimismo, es posible lograr la ordenación de los datos en dos o tres dimensiones, apoyando la sistematización de las OTUs y una mejor visualización de las relaciones entre ellas (Crisci y López 1983; De Vicente *et al.*, 2004).

Con los marcadores genéticos moleculares se pueden examinar un grupo de individuos o poblaciones para estimar medidas de diversidad y distancias genéticas, para inferir estructuras poblacionales y patrones de agrupamiento, comprobar equilibrios de Hardy Weinberg y de multiloci y buscar en loci polimórficos la evidencia de neutralidad de selección Sokal y Sneath (1963). Asimismo, existen una amplia gama disponible de programas informáticos de enfoque estadístico para estudios moleculares y de genética de poblaciones, y destinados para computadoras personales. Estos pueden acomodar una variedad de marcadores moleculares y llevar a cabo diferentes tipos de análisis (Yeh *et al.*, 1999; Rohlf, 2000; Excoffier *et al.*, 2005).

### 3. HIPÓTESIS Y OBJETIVOS

Se plantearon las siguientes hipótesis:

- La diversidad genética que albergan los individuos (clones e híbridos) a analizar es elevada.
- Hay una concordancia de los agrupamientos moleculares con la clasificación clásica del cacao por grupos raciales: Criollo, Forastero y Trinitario.
- Existen marcadores ISSR comunes entre los árboles híbridos y sus respectivos clones parentales.
- Se encuentran marcadores ISSR asociados con rasgos de interés agronómico.

Se propusieron los siguientes objetivos:

- Evaluar la potencialidad de los ISSR para determinar el grado de similaridad molecular entre los individuos de la colección de cacao en estudio.
- Investigar si los agrupamientos generados se relacionan directamente con caracteres morfológicos y agronómicos recopilados de la colección estudiada.
- Establecer las relaciones intrapoblacionales e interpoblacionales entre 44 individuos de cacao agrupados *a priori* según origen geográfico.

## 4. MATERIALES Y MÉTODOS

### 4.1 Ámbito de estudio.

Los 50 individuos de cacao, clones o progenie híbrida, 30 clones nacionales, 14 internacionales y 6 híbridos, provenientes de diferentes regiones de Sudamérica, Centroamérica o el Caribe. En la Tabla N° 4 se muestran los datos por individuo.

**Tabla N° 4:** Lista de individuos de cacao utilizados para el estudio con marcadores ISSR. Parte A: Clones nacionales, Parte B: Clones internacionales, Parte C: Híbridos Bioversity/UNAS.

Parte A: Clones Nacionales			
n°	código	Grupo o raza	Origen Geográfico
1	C85	Forastero x Trinitario	Echarate, Cusco
2	ChuC	Forastero alto amazónico	Echarate, Cusco
3	Seño	Forastero alto amazónico	Echarate, Cusco
4	SeñoA	Forastero alto amazónico	Echarate, Cusco
5	U54	Forastero alto amazónico	Quillabamba, Cusco
6	U60	Forastero alto amazónico	Quillabamba, Cusco
7	U68	Forastero alto amazónico	Quillabamba, Cusco
8	U70	Forastero alto amazónico	Quillabamba, Cusco
9	H12	Forastero x Trinitario	Tingo María, Huánuco
10	H37	Forastero x Trinitario	Tingo María, Huánuco
11	H60	Forastero x Trinitario	Alomia Robles, Huánuco
12	H61	Forastero x Trinitario	Alomia Robles, Huánuco
13	SilvHu	Forastero alto amazónico	Alomia Robles, Huánuco
14	ChuS	Forastero alto amazónico	Satipo, Junin
15	IMC67	Forastero alto amazónico	Iquitos, Loreto
16	P7	Forastero alto amazónico	Iquitos, Loreto
17	SCA6	Forastero alto amazónico	Iquitos Loreto
18	PA150	Forastero alto amazónico	Parinari, Loreto
19	C42	Forastero x Trinitario	Palo de Acero, San Martín
20	C53	Forastero x Trinitario	Pumahuasi, San Martín
21	C65	Forastero x Trinitario	San Alejandro, San Martín
22	C69	Forastero x Trinitario	Tocache, San Martín
23	C81	Forastero x Trinitario	Nuevo Horizonte, San Martín
24	U1	Forastero alto amazónico	Cuenca del río Ucayali
25	U9	Forastero alto amazónico	Cuenca del río Ucayali
26	U12	Forastero alto amazónico	Cuenca del río Ucayali
27	U15	Forastero alto amazónico	Cuenca del río Ucayali
28	U26	Forastero alto amazónico	Cuenca del río Ucayali
29	U43	Forastero alto amazónico	Cuenca del río Ucayali
30	U48	Forastero alto amazónico	Cuenca del río Ucayali

Parte B: Clones Internacionales			
nº	código	Grupo o raza	Origen Geografico
1	CAT4	Forastero bajo amazónico	Brasil
2	Pand	desconocido	Colombia
3	CCN51	Complejo x Forastero	Ecuador
4	EET228	híbrido amazónico	Ecuador
5	EET233	híbrido	Ecuador
6	EET400	Forastero alto amazónico	Ecuador
7	EET62	Nacional x desconocido	Ecuador
8	ICS1	Trinitario	Trinidad y Tobago
9	ICS6	Trinitario	Trinidad y Tobago
10	ICS39	Trinitario	Trinidad y Tobago
11	ICS78	Trinitario	Trinidad y Tobago
12	ICS95	Trinitario	Trinidad y Tobago
13	UF29	Forastero x Nacional	Costa Rica
14	UF667	Trinitario	Costa Rica

Parte C: Híbridos Proyecto Bioversity/UNAS				
nº	código	Parentales	Grupo o raza	Origen Geografico
1	HTM1	CCN51 x EET233	Complejo	no aplicable
2	I12.12	IMC67 x EET228	Trinitario	no aplicable
3	I14.20	P7 x ICS95	Forastero x Trinitario	no aplicable
4	I16.20	ICS1 x SCA6	Trinitario x Forastero	no aplicable
5	M1.7	H12 x ICS6	Forastero x Trinitario	no aplicable
6	M18.16	IMC67 x U68	híbrido Forastero	no aplicable

**Fuentes:** Soria y Enriquez, (Eds.) 1981; Bartra, 1993; N'Goran *et al.*, 1994; Gonzales, 1996; Marita *et al.*, 2001; Rengifo, 2002; Guerrero, 2005; Sounigo *et al.*, 2005; García, 2007; García, 2008 (com. per.).

**Observaciones:** **ChuC** significa "Chuncho" del Cusco; **Seño** es la abreviatura de cacao "Señorita"; **SeñoA** es la abreviatura de cacao "Señorita Achoccha"; **SilvHu** es la abreviatura de cacao "silvestre Huallaga"; **P7** es la abreviatura de cacao "Pound 7"; el genotipo "Pandora" se abrevió como el código **Pand**; el genotipo "Catongo 4" se abrevió **CAT4** y **HTM1** es la abreviatura de "Híbrido Tingo Maria 1".

## 4.2 Material vegetal

Los clones nacionales e internacionales mostrados en la Tabla N° 4 parte A y B se ubican en el Banco de germoplasma de cacao de la UNAS y los híbridos de la parte C se localizan en la E.E.A. Tulumayo; ambos lugares situados en la provincia de Leoncio Prado, región de Huánuco. Todas las accesiones se propagan vegetativamente, por lo tanto se eligió uno o dos árboles clonales y se colectaron hojas jóvenes sin lesiones por insectos, ni infectadas gravemente con patógenos. Fueron envueltas en papel toalla

limpio, se marcó su código respectivo con plumón indeleble y se almacenaron en una bolsa plástica conteniendo silicagel deshidratado previamente. Los detalles morfológicos y características agronómicas de importancia se muestran en los Anexos N° 1 y N° 2. Estos datos fueron recopilados con la finalidad de usarse para la interpretación y discusión de los resultados moleculares.

#### **4.3 Aislamiento y purificación del ADN**

Se realizaron pruebas preliminares de aislamiento y purificación de ADN de acuerdo a Doyle y Doyle (1987), con muestras de hoja de cuatro accesiones con código: ICS1, CAT4, U70 y EET223. Estos ensayos se realizaron con la finalidad de asegurar que la metodología a usar rendía buenos resultados a nivel de pureza, calidad y cantidad del ADN extraído. Se pesó la cantidad aproximada de 100 mg de tejido (completamente deshidratados con silicagel) para cada una de las muestras. Cada una de éstas se colocaron en un microtubo nuevo y estéril de 1.5 ml junto con unas bolitas de acero estériles para ser molidas durante 5 minutos a temperatura ambiente en un *Bead Mill Retsch MM200* hasta obtener un polvo fino. A cada microtubo conteniendo el tejido molido se agregaron 800 µl de buffer CTAB y se mezcló suavemente por inversión hasta obtener una suspensión homogénea. Los microtubos se colocaron en baño maría a 60 °C por 30 minutos, agitándolos suavemente cada 10 minutos. Las muestras se dejaron enfriar y se agregaron 800 µl de cloroformo - alcohol isoamílico en una proporción de 24:1. Cada muestra se mezcló por inversión hasta tener un solución lechosa y luego se centrifugó a 14 000 r.p.m. por 10 minutos. El sobrenadante se transfirió cuidadosamente a un nuevo tubo con una micropipeta adecuada. Agregando 800 µl de etanol al 96% y mezclando cuidadosamente por inversión, se consigue precipitar el ADN como una masa flotante o medusa. La muestra se centrifugó por un minuto a 10 000 r.p.m. para obtener el ADN como un pellet. El alcohol se eliminó con mucho cuidado. Se realizaron dos lavados del pellet con etanol al 70%, a temperatura ambiente por 5 minutos, centrifugando finalmente a 10 000 r.p.m. por 5 minutos. Se eliminó el sobrenadante y los microtubos se



dejaron con la tapa abierta e invertidos dentro de una cabina de flujo laminar durante toda la noche. Posteriormente, se resuspendió el precipitado de ADN en 50  $\mu$ l de TE y se procedió a verificar la calidad de ADN en gel de agarosa y la pureza en el biofotómetro.

#### **4.4 Calidad del ADN**

La evaluación de la calidad e integridad el ADN se llevó a cabo en electroforesis en gel de agarosa. La concentración fue de 1% en buffer TBE. La agarosa se llevó a microondas hasta casi hervir, se dejó enfriar hasta aproximadamente 60 °C y luego se agregaron 2  $\mu$ l de bromuro de etidio (2mg/ml) para 100 ml de buffer. Se vertió en el molde con los peines y se esperó hasta que gelifique. El gel listo fue colocado en la cámara de electroforesis conteniendo buffer TBE 1X. En los pocillos del gel se cargaron 2  $\mu$ l de la muestra previamente mezclados con 8  $\mu$ l de buffer de carga. Se utilizó un ADN control y las muestras de los cuatro genotipos se corrieron por duplicado. Se aplicó corriente a 80 voltios por una hora, transcurrido este tiempo se visualizó y capturó la imagen del ADN en el *BioRad Gel Doc 2000*. Se muestra en la Figura N° 9.

#### **4.5 Determinación de la Pureza y Concentración del ADN**

Cada muestra de ADN extraído fue sometida a incubación con 2  $\mu$ l de ribonucleasa a 37 °C por dos horas. Se realizaron las lecturas en un biofotómetro Eppendorf, el mismo que proporcionó directamente la relación  $A_{260}/A_{280}$  (tasa de absorbancias a 260 nm sobre 280 nm) considerada como pureza del ADN entre los valores de 1,8 a 2,0. Asimismo, proporcionó directamente la concentración de la muestra (ng/ $\mu$ l) a partir de la fórmula empírica:  $50 \times A_{260} \times \text{Factor de dilución}$ . Confirmada la adecuada pureza y la concentración de ADN para estos 4 individuos, se inició la extracción y purificación del ADN de los 46 individuos restantes. En el Anexo N° 3 se muestran los datos de pureza ( $A_{260}/A_{280}$ ) y concentraciones de ADN (ng/ $\mu$ l) para todas las accesiones.

#### **4.6 Pre-selección de iniciadores ISSR**

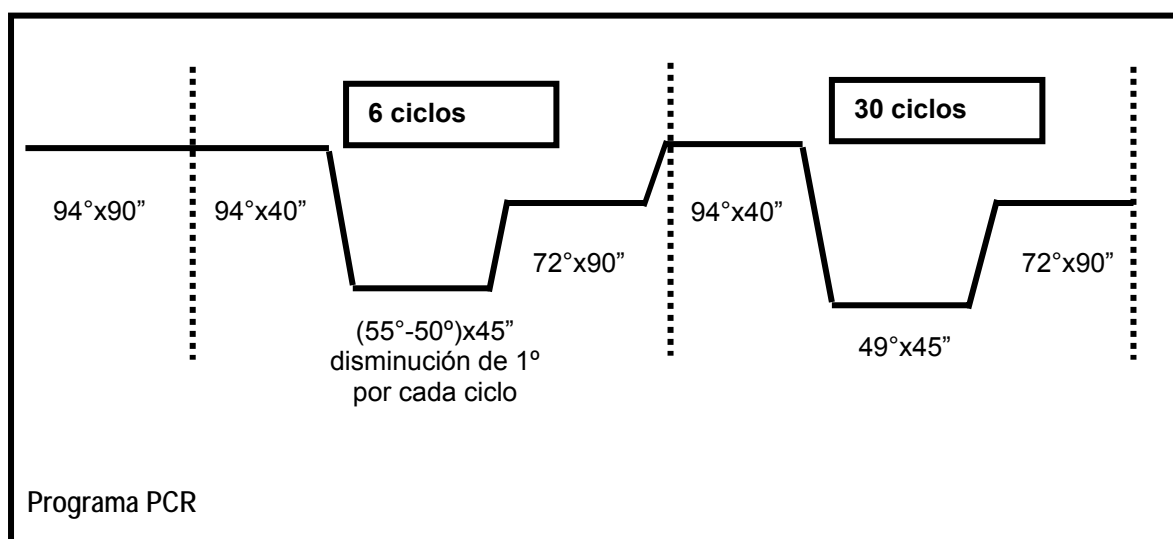
Con la finalidad de elegir los iniciadores que proporcionen una mejor respuesta en términos de amplificación de bandas nítidas y polimórficas; se realizó una preselección de 59 iniciadores ISSR (Anexo N° 4). Se amplificaron los ADN de 4 accesiones de cacao: ICS1, CAT4, U70 y EET223, cada uno con su duplicado de ADN aislado independientemente (réplica biológica). Se comparó el patrón de bandas para verificar la reproducibilidad de éstas.

#### **4.7 PCR con los 50 individuos del estudio**

A partir de los resultados de la pre-selección del apartado anterior, se seleccionó un grupo de iniciadores con los cuales se realizó la amplificación del ADN de las 50 accesiones (previamente aisladas y purificadas de acuerdo a lo expuesto en el subtítulo 3.3).

#### **4.8 Condiciones PCR – ISSR**

Para realizar las reacciones en cadena de la polimerasa (PCR) se procedieron a diluir las muestras de ADN hasta una concentración de 25 ng/μl. Se llevó a cabo en un volumen final de 20 μl, usando las siguientes cantidades por tubo de reacción: 15 ng de ADN; 1X PCR buffer (10 mM TRIS-HCl, pH 8.3, 50 mM KCl); 0.2 mM dNTPs; 3 mM MgCl<sub>2</sub> ; 0.8 μM iniciador y 1 U *Taq* polimerasa (Fermentas). Se llevaron a cabo en un termociclador Eppendorf *Mastercycle gradient* 96. El programa de amplificación se muestra esquemáticamente en la Figura N° 7.

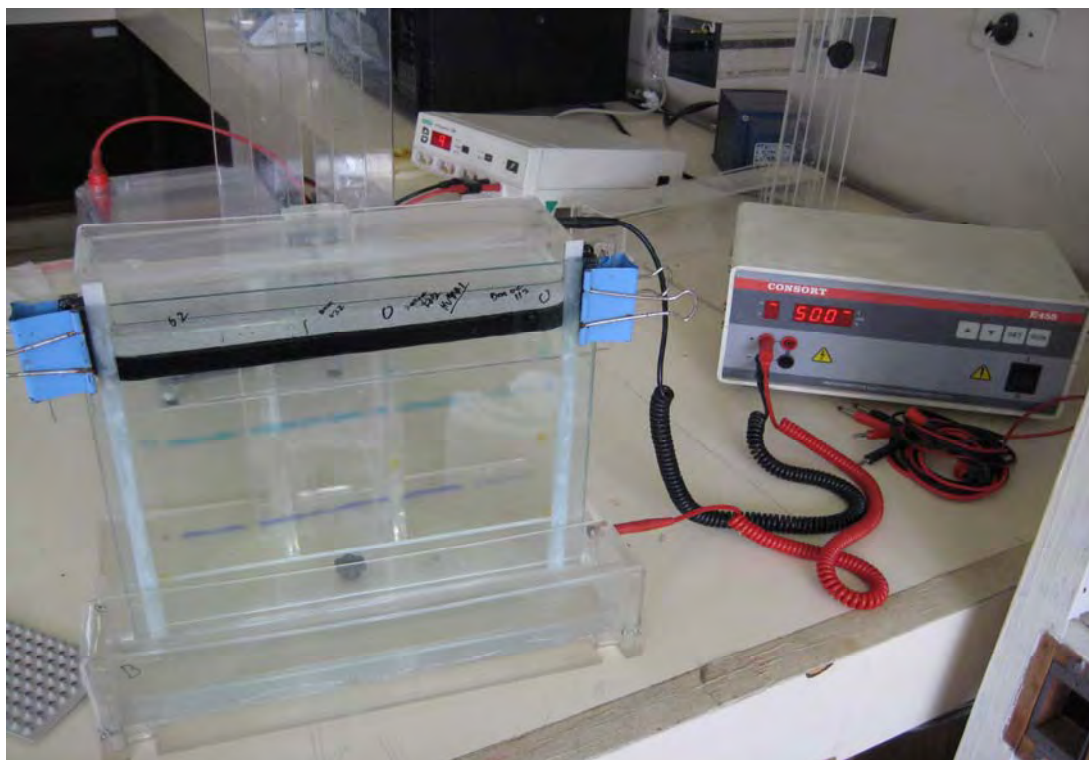


**Figura N° 7:** Esquema del programa PCR-ISSR en el termociclador *Eppendorf Mastercycle gradient 96*.

#### 4.9 Electroforesis en gel denaturante de poliacrilamida (PAGE)

La electroforesis se realizó en cámaras verticales (Figura N° 8) diseñadas y construidas en base a modelos del Instituto Max Planck de Alemania, que además proporcionó los materiales para su construcción. El gel de poliacrilamida se preparó al 6% en una proporción de 19:1 (acrilamida/bisacrilamida) y 7 M de úrea, catalizada con 40 µl de TEMED y 400 µl de persulfato de amonio al 10%. La cámara electroforética contiene 800 ml de tampón Tris–borato EDTA (TBE) 1x. Antes de cargar las muestras, se realizó una pre-corrida sin muestras a 500 V por 30 minutos y posteriormente se hace la corrida con la muestras a 500 V por 6 horas. A cada amplificación se añadió 10 µl de tampón de carga conteniendo 95% de formamida, 20 mM de EDTA, 0.05 µl de azul de bromofenol y 0.05 µl de xilen cianol. Se denaturalizó cada muestra a 94 °C por 3 minutos y se cargó en sus respectivos pocillos. Se corrió a ambos lados del gel un marcador de peso molecular: *O'GeneRuler 100 bp DNA Ladder Plus* (Fermentas). La fijación se realizó con 100 ml de etanol y 5 ml de ácido acético para un litro de solución. Esto se dejó por 20 minutos. El buffer de tinción fue el mismo que la solución de fijación al cual se le añadió 2 g de nitrato de plata. El gel

se dejó en esta solución por 20 minutos adicionales. El revelado se realizó utilizando 30 g de hidróxido de sodio en un litro de agua y 2.1 ml de formaldehído. Dependió del tiempo la aparición de las bandas. Luego de esto, el gel fue puesto nuevamente por 5 minutos en la solución de fijación para detener la reacción, y finalmente se realizó un enjuague en agua corriente por 3 ó 4 minutos. El gel se dejó secar al ambiente.



**Figura N° 8:** Fotografía de una cámara vertical de electroforesis en poliacrilamida construida artesanalmente. Se muestra junto con la fuente de poder.

#### **4.10 Colecta de los datos moleculares y determinación de valores descriptivos**

Se usó el programa Cross Checker v2.9 (Buntjer, 1999) para la asignación de los valores binarios de presencia (1) y ausencia (0) de fragmentos. Visualizando las fotografías digitales de los PAGE, se trazaron líneas para los carriles y para los loci, en la intersección de ambos se asigna el valor correspondiente si hubiera presencia o ausencia de banda de ADN. A los fragmentos con tenue amplificación, se les asignó un valor de 2,

consignándolos como datos perdidos o *missing data*. Esta información de bandas se ingresó en hojas de cálculo de Microsoft Excel 2000 (Anexo N° 5). En esta misma hoja de cálculo, se determinó el porcentaje de datos perdidos por cada locus ingresado, de la siguiente forma: la cantidad de datos perdidos por locus se dividió entre el número total de OTUs, en este caso cincuenta, y se multiplicó por cien. Los loci con un porcentaje de datos perdidos mayor al 10 % fueron descartados para análisis posteriores (Warburton y Crossa, 2002).

Se calcularon las frecuencias de presencias ( $p$ ) para cada loci. Se dividió la cantidad de fragmentos amplificados (1) entre el total de individuos sin considerar los datos perdidos. La frecuencia de ausencias ( $q$ ) se obtuvo restando 1 menos el valor de  $p$ . Los loci con una frecuencia de presencia de banda ( $p$ ) menores a 0.05 se consideraron alelos raros. Los loci con una frecuencia de presencia de banda ( $p$ ) igual a 1.0 son considerados monomórficos. Para conocer la utilidad de los marcadores moleculares, se cuentan con tres medidas: riqueza de alelos, *Polymorphic Index Content* (PIC) y el poder discriminatorio de los marcadores. Para este trabajo se calculó el PIC para cada locus con la siguiente fórmula (Ghislain *et al.*, 1999):

$$PIC = 1 - \sum_{i=1}^n f_i^2 = 1 - p^2 - q^2$$

$f_i$ : frecuencia de banda del  $i$ -ésimo alelo.

$p$ : frecuencia de presencia de banda.

$q$ : frecuencia de ausencia de banda.

Finalmente, se sumaron todos los valores PIC para cada iniciador ISSR, generando sus respectivos índices informativos ISSR (basado en Ghislain *et al.*, 1999). Todos estos valores descriptivos se muestran en el Anexo N° 6.

#### 4.11 Análisis Multivariado

Con la finalidad de entender la diversidad molecular entre todos los miembros de la colección estudiada se realizó un análisis multivariado de 73 loci (caracteres) y 50 OTUs (*taxa*), ordenados en una “matriz básica”, siendo las filas los loci y las columnas las OTUs (Anexo N° 5). Se utilizó el paquete estadístico NTSYS v2.1p (Rohlf, 2000). Mediante este software se realizaron tres análisis: agrupamiento por UPGMA, ordenación mediante Análisis de Coordenadas Principales (en sus siglas en inglés: PcoA, *Principal Coordinate Analysis*) y ordenación por Escalamiento Multidimensional no métrico (en sus siglas en inglés: nmMDS, *nonmetric Multidimensional Scaling*). Paralelamente, con la finalidad de determinar si se establecían relaciones de similitud entre híbridos y sus respectivos parentales, a un grupo de 17 accesiones, consistente en 6 híbridos Bioversity/UNAS y 11 respectivos progenitores se realizaron análisis multivariados similares a los mencionados

En primer lugar, se calculó la similitud entre cada par de genotipos usando los métodos para calcular la diversidad basada en diferencias cualitativas. Gracias a esto se logra la construcción de una “Matriz de Similitud”. En la comparación de OTU<sub>i</sub> vs. OTU<sub>j</sub>, se pueden dar las siguientes cuatro posibilidades (Crisci y López, 1983): coincidencia del carácter presencia (1,1) en ambos genotipos, presencia en el genotipo *i* y no en el *j* (1,0); ausencia en el genotipo *i* y no en el *j* (0,1) y coincidencia de ausencias en ambos genotipos (0,0). Esto se resume en la siguiente matriz:

		OTU <sub>j</sub>	
		1	0
OTU <sub>i</sub>	1	a	b
	0	c	d

Para obtener la matriz de similitud se recurren a diferentes coeficientes de asociación (o coeficientes de similitud), de los cuales dos son los más

utilizados con datos moleculares: coeficiente de *Jaccard* y coeficiente de *Dice* (Crisci y López, 1983; Rohlf, 2000). Estos coeficientes no consideran a d(0,0) como elemento a favor de la similitud.

$$Jaccard = \frac{a_{(1,1)}}{a_{(1,1)} + b_{(1,0)} + c_{(0,1)}}$$

$$Dice = \frac{2a_{(1,1)}}{2a_{(1,1)} + b_{(1,0)} + c_{(0,1)}}$$

El coeficiente *Dice* llamado también de Sorensen o Nei-Li (Crisci y López, 1983; De Vicente *et al.*, 2004) confiere peso duplicado a los valores de bandas compartidas entre ambos genotipos. Los valores de similitud generados por ambos coeficientes varían desde 0 hasta 1, que equivalen a los valores de mínima y máxima similaridad, respectivamente.

Entonces, se ordenaron estas comparaciones uno a uno entre las OTUs, y se generó una Matriz de Similitud (Anexo N° 9). Obtenida ésta, se llevó a cabo un agrupamiento UPGMA (*Unweighted Pair Group Method using Arithmetic Averages*), que es un algoritmo jerárquico que agrupa a los taxones en grupos similares, teniendo en cuenta las distancias geométricas entre OTUs. El resultado de este procedimiento se grafica en un dendograma o árbol fenético. Se dibujaron dos árboles, utilizando ambos coeficientes de asociación mencionados y se compararon con la finalidad de definir cuál es el que explicaría mejor la diversidad molecular entre las accesiones.

Con la finalidad de comprobar la independencia de ambas matrices y el nivel de confianza, se realizó una validación interna (De Vicente *et al.*, 2004), mediante la comparación de las matrices de distancia y cofenética, y se determinó el coeficiente de correlación cofenética o de Mantel (*r*). Esta es conocida como Prueba de Mantel (Rohlf, 2000).

Finalmente, para una disposición o “arreglo” de las OTUs en un sistema de coordenadas con la finalidad de formar e interpretar patrones en la

composición de las OTUs, se representaron éstas en dos dimensiones mediante el Análisis de Coordenadas Principales y el Análisis de Escalamiento Multidimensional no métrico, basándose con el mejor coeficiente de Asociación determinado en la comparación realizada en base a los dendogramas. Se generaron valores de *FinalSTRESS2*, para el caso del nmMDS, los cuales son indicadores de bondad de ajuste o “*Goodness of fit*” del ordenamiento de datos. Tales valores se califican dependiendo de la siguiente escala (Rohlf, 2000):

Final Stress2	Calificativo del nmMDS
0.40	Pobre
0.20	Mediano
0.10	Bueno
0.05	Excelente
0.00	"Perfecto"

En el paquete NTSYSpc v2.1p, los comandos que se ejecutaron para los análisis multivariados descritos anteriormente se escribieron en un “archivo lote” o *batch file*. Las líneas escritas a partir de las comillas (“) son comentarios y no son leídos por el software; en cambio, las líneas escritas a partir del asterisco (\*) son los comandos junto con sus parámetros correspondientes (Rohlf, 2000). Estos *batch files* se presentan en el Anexo N° 8.

#### 4.12 Análisis Poblacional

Para el estudio de la diversidad genética se investigó la variación presente por medio de una comparación jerárquica entre poblaciones y subpoblaciones. Se usaron los softwares genéticos: Arlequin v3.1 (Excoffier *et al.*, 2005), POPGENE v1.32 (Yeh *et al.*, 1999) y Treeview v1.6 (Page, 2000).



Los clones de cacao se dividieron en dos poblaciones: Nacionales e Internacionales (Tabla N° 4). A su vez, la primera población se dividió en 6 subpoblaciones según el departamento de origen: Cusco, Huánuco, Junín, Loreto, San Martín y Ucayali. De igual forma, la población internacional se particionó en sus respectivos países de origen: Brasil, Ecuador, Costa Rica, Colombia y Trinidad y Tobago. Se excluyeron de este análisis los híbridos Bioversity/UNAS, dado que no se les puede asignar un origen geográfico definido ya que provienen de cacao parentales de diferentes regiones.

En el programa POPGENE no fue posible especificar el coeficiente de endogamia ( $F_{IS}$ ), para cada subpoblación, ya que este parámetro debe ser calculado previamente con marcadores codominantes. Por ello, para lograr el cálculo de estos parámetros, se debió hacer la suposición de que la población está en un equilibrio Hardy Weinberg. Mediante este software se estimaron los siguientes parámetros:

- Loci polimórficos para población, subpoblación y todos los individuos.
- Índice de Información de Shannon como medida de la diversidad genética:

$$I = - \sum_{i=1}^n p_i^2 \log_2 p_i$$

- Diversidad Genética de Nei, conocida como Heterocigocidad esperada ( $H_e$ ) y es la probabilidad de que cualquier par de alelos (en un único locus) escogidos al azar de la población sean diferentes entre si (De Vicente *et al.*, 2004).
- Parámetro  $G_{st}$  entre poblaciones y entre pares de subpoblaciones.
- Matriz de Distancias e Identidades Genéticas normalizadas (unbiased) de Nei (1978) (Yeh, *et al.*, 1999).

Con el software Treeview se visualizó el Dendograma sin raíz de las subpoblaciones, generado en el programa POPGENE.

Por último, se cuantificó la variación presente entre los componentes individuos, poblaciones y subpoblaciones, mediante el AMOVA (Análisis Molecular de Variancia), usando el programa Arlequin v3.1 con la opción “*pairwise differences*”. En este software, los datos moleculares binarios se consignaron como haplotipos de RFLP (cuyos loci son sitios completamente ligados sin ninguna recombinación) a diferencia de los generados por ISSR o marcadores dominantes (sitios independientes con equilibrio de ligamiento). Por ello el índice  $F_{st}$  (adecuadamente se debe simbolizar como  $\Phi_{st}$  para marcadores dominantes) no se puede comparar con el índice obtenido de marcadores codominantes (De Vicente *et. al.*, 2004).

## 5. RESULTADOS

### 5.1 Pruebas de aislamiento y purificación del ADN.

El protocolo de aislamiento y purificación del ADN utilizado de acuerdo a Doyle y Doyle (1987), proporcionó resultados aceptables en términos de pureza y cantidad de ADN. En la Tabla N° 5, se aprecian los valores de pureza (relación  $A_{260}/A_{280}$ ) y concentraciones de ADN en ng/μl para los 4 genotipos ensayados preliminarmente. La pureza y la concentración de ADN para los 50 individuos se muestran en el Anexo N° 3. Ahí se pueden observar que los valores de pureza entre las 50 accesiones varían desde 1.20 hasta 2.39 y desde 111 hasta 1802 ng/μl en concentración.

**Tabla N° 5:** Índices de pureza y concentraciones de ADN obtenidas en lecturas con biofotómetro para 4 accesiones de cacao.

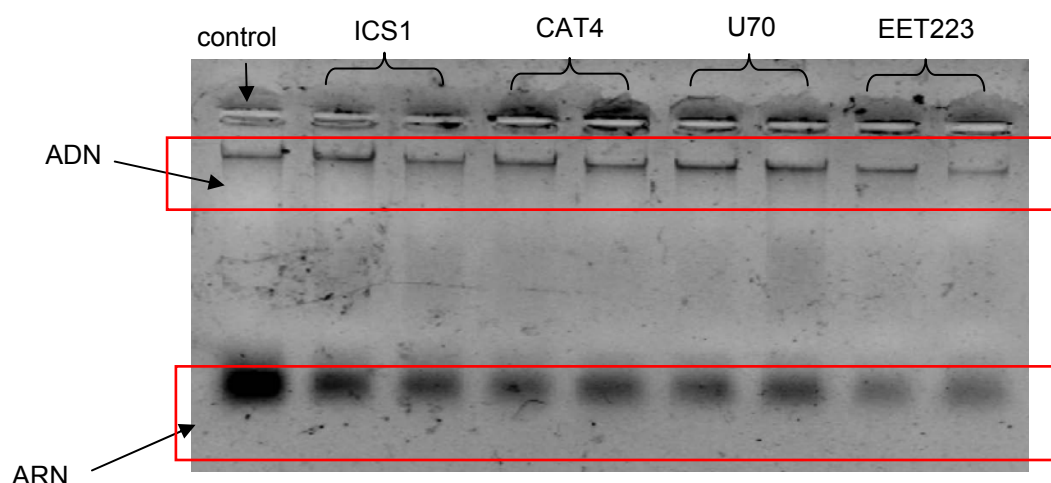
Código Accesiones	Relación $A_{260}/A_{280}$	Concentración ADN (ng/μl)
ICS1	1,56	630
CAT4	1,43	301
U70	1,31	139
EET223	1,29	449

### 5.2 Calidad del ADN

En la Figura N° 9 se muestra el gel de agarosa con las cuatro muestras de ADN aisladas (en duplicado) y un control. Una buena calidad se refiere a una mayor integridad de las moléculas de ADN genómico. En esta misma Figura, se visualizaron bandas de ARN en las 4 muestras, por lo que se recurrió a la ribonucleasa después de la purificación del ADN. La digestión con esta enzima se realizó posteriormente con todas las 50 muestras de ADN, y se comprobó su efectividad mediante las lecturas espectrofotométricas (Anexo N° 3).

### 5.3 Elección de los iniciadores.

Las pruebas de amplificación de los 59 iniciadores ISSR con las 4 accesiones de prueba (cada una con su réplica biológica); lograron determinar una lista de iniciadores cuyo patrón de bandas fue satisfactorio y reproducible. Asimismo, resultaron con una considerable cantidad de bandas polimórficas (Tabla N° 6).



**Figura N° 9:** Electroforesis en gel de agarosa al 1% mostrando la calidad de ADN aislado por duplicado para 4 accesiones de cacao. Se indica un control en el primer carril.

**Tabla N° 6:** Lista de los 13 Iniciadores ISSR que dieron buena respuesta en la amplificación de 4 muestras (por duplicado) de ADN de cacao.

Iniciador	Bandas totales	Bandas Polimórficas
Gup1	15	7
Gup2	35	9
Bor5	26	5
UBC829	12	10
UBC830	10	4
UBC834	29	12
UBC835	22	20
UBC841	31	10
UBC842	30	8
UBC864	23	9
UBC868	15	6
UBC880	20	6
UBC884	30	7

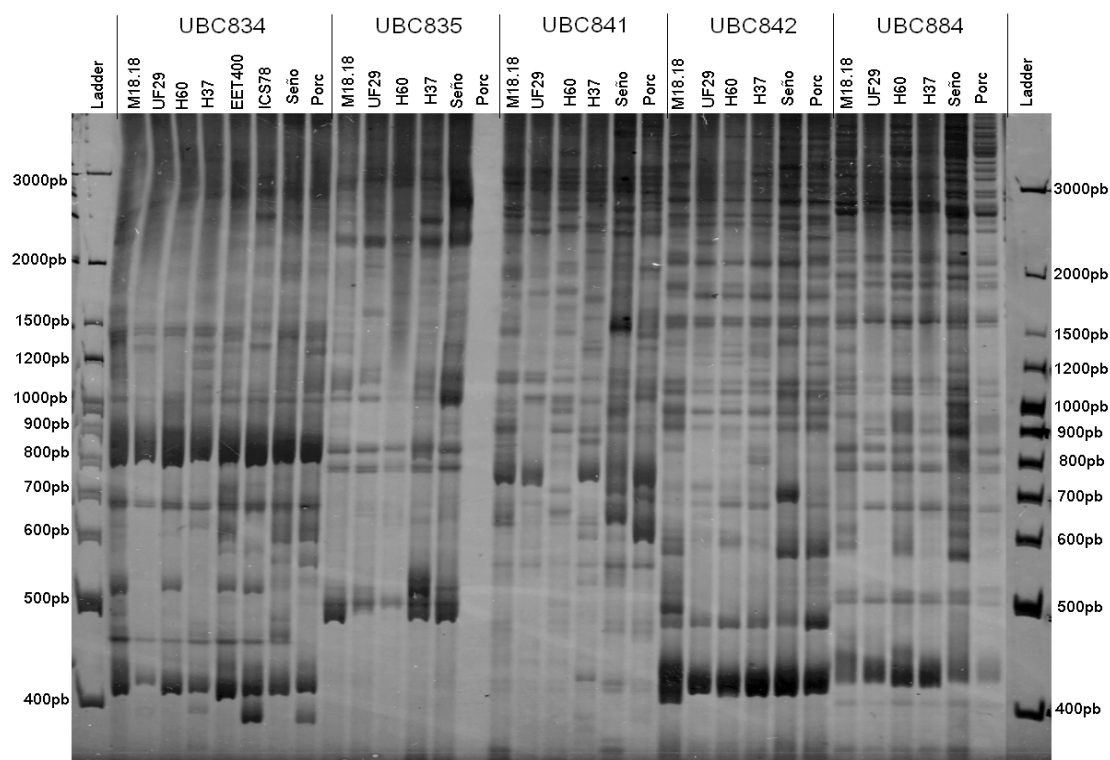
Con el resto de iniciadores no se logró una buena amplificación (Lista de códigos no resaltado en el Anexo N° 4); sin embargo, no se pueden descartar en un 100%, ya que no se realizaron adicionales pruebas de amplificación PCR.

#### 5.4 Ensayos ISSR de la colección completa.

Las amplificaciones del ADN con todos los individuos se realizaron eligiendo 5 de los iniciadores ISSR que dieron mejor respuesta en el proceso de pre-selección (Tabla N° 7). Se muestran la cantidad de loci totales amplificados por iniciador; además la cantidad de loci polimórficos, el “Índice de Iniciador o Índice ISSR” y el número de fragmentos raros (en el Anexo N° 6 se muestra en detalle el cálculo de tales valores). Estos valores fueron obtenidos después del proceso de colección de datos y exclusión de los datos perdidos (como se describió en el subtítulo 3.10 de la metodología). En la Figura N° 10, se muestra una corrida electroforética en gel de poliacrilamida de las amplificaciones con los 5 iniciadores ISSR utilizados en esta tesis y un grupo de los genotipos de cacao. Se puede observar que hay un considerable polimorfismo entre las amplificaciones.

**Tabla N° 7:** Lista de Iniciadores ISSR utilizados para el estudio molecular de las 50 accesiones de cacao. El índice ISSR es la suma de todos los PIC de los loci para cada iniciador. La cantidad de fragmentos raros ( $p \leq 0,05$ ) es por genotipo encontrado.

Código	Secuencia	N° loci	N° loci polimórficos	Índice ISSR	N° fragmentos raros
UBC834	(AG) <sub>8</sub> YT	11	7	2,07	1
UBC835	(AG) <sub>8</sub> YC	14	13	1,63	5
UBC841	(GA) <sub>8</sub> YC	19	19	2,43	16
UBC842	(GA) <sub>8</sub> YG	16	15	4,56	8
UBC884	HBH(AG) <sub>7</sub>	13	12	4,20	2

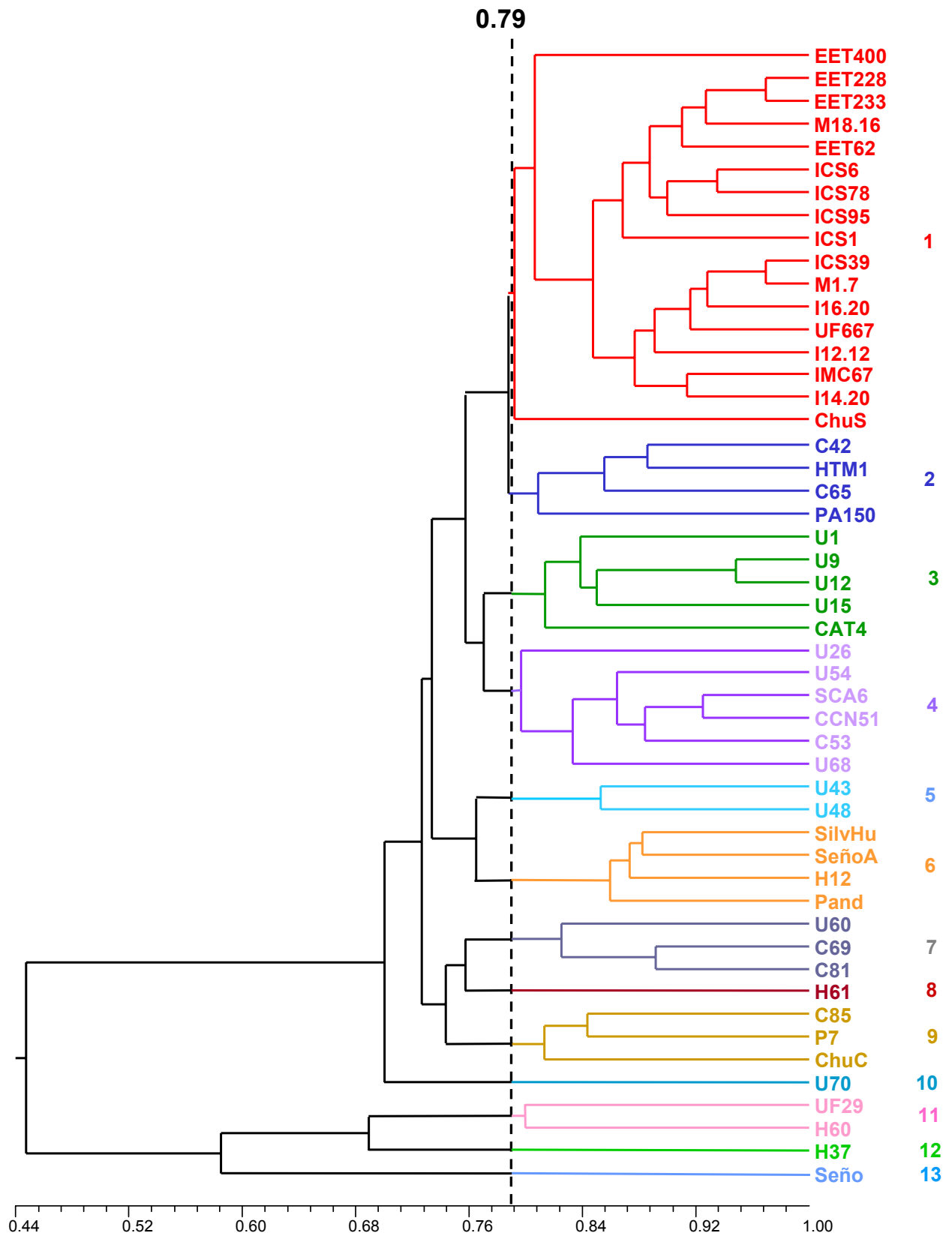


**Figura N° 10:** Electroforesis en gel de poliacrilamida al 6% mostrando los patrones de amplificación de 5 *primers* ISSR usados (UBC834, UBC835, UBC841, UBC842 y UBC884) con los genotipos cuyos códigos se muestran en el lado superior del carril. Marcador de peso molecular en ambos extremos: O'GeneRuler 100 bp DNA Ladder Plus (Fermentas).

## 5.5 Análisis multivariado

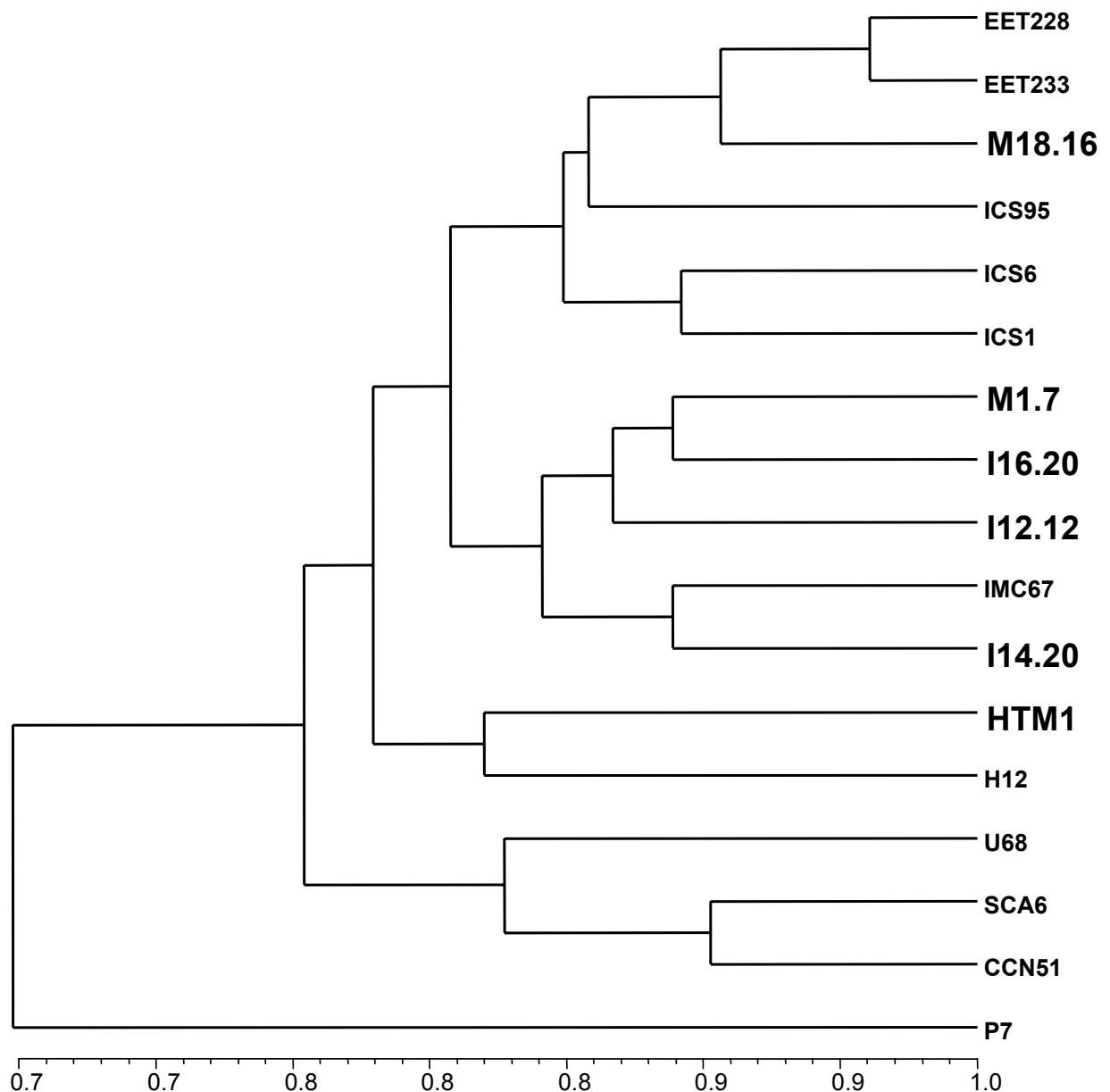
### 5.5.1 Agrupamiento UPGMA

Usando el programa NTSYSpc v2.1p se generaron fenogramas para las 50 accesiones de cacao con dos coeficientes de asociación o similitud (*Jaccard* y *Dice*). Al comparar ambos árboles, se eligió aquel generado con *Jaccard*. Éste se muestra en la Figura N° 11. Se escogió un valor de similitud de 0.79 para resolver 13 agrupamientos o “*clusters*”, los cuales se incluyen en la Figura N° 11 con diferentes colores.



**Figura N° 11:** Fenograma generado con el coeficiente de asociación *Jaccard*. A un valor de 0,79 de similitud, se muestran 13 *clusters* en diferentes colores. El coeficiente de correlación cofenética ( $r$ ) es de 0,91383.

El fenograma con el coeficiente de similitud *Dice*, se muestra en el Anexo N° 10. Las figuras de correlación entre matrices de distancia y cofenético con el coeficiente *Jaccard* se muestran en el Anexo N° 11. Asimismo, se realizó el agrupamiento de los 6 Híbridos Bioversity/UNAS con sus respectivos parentales (Tabla N° 4) con el coeficiente *Jaccard*. El fenograma generado se muestra en la Figura N° 12.



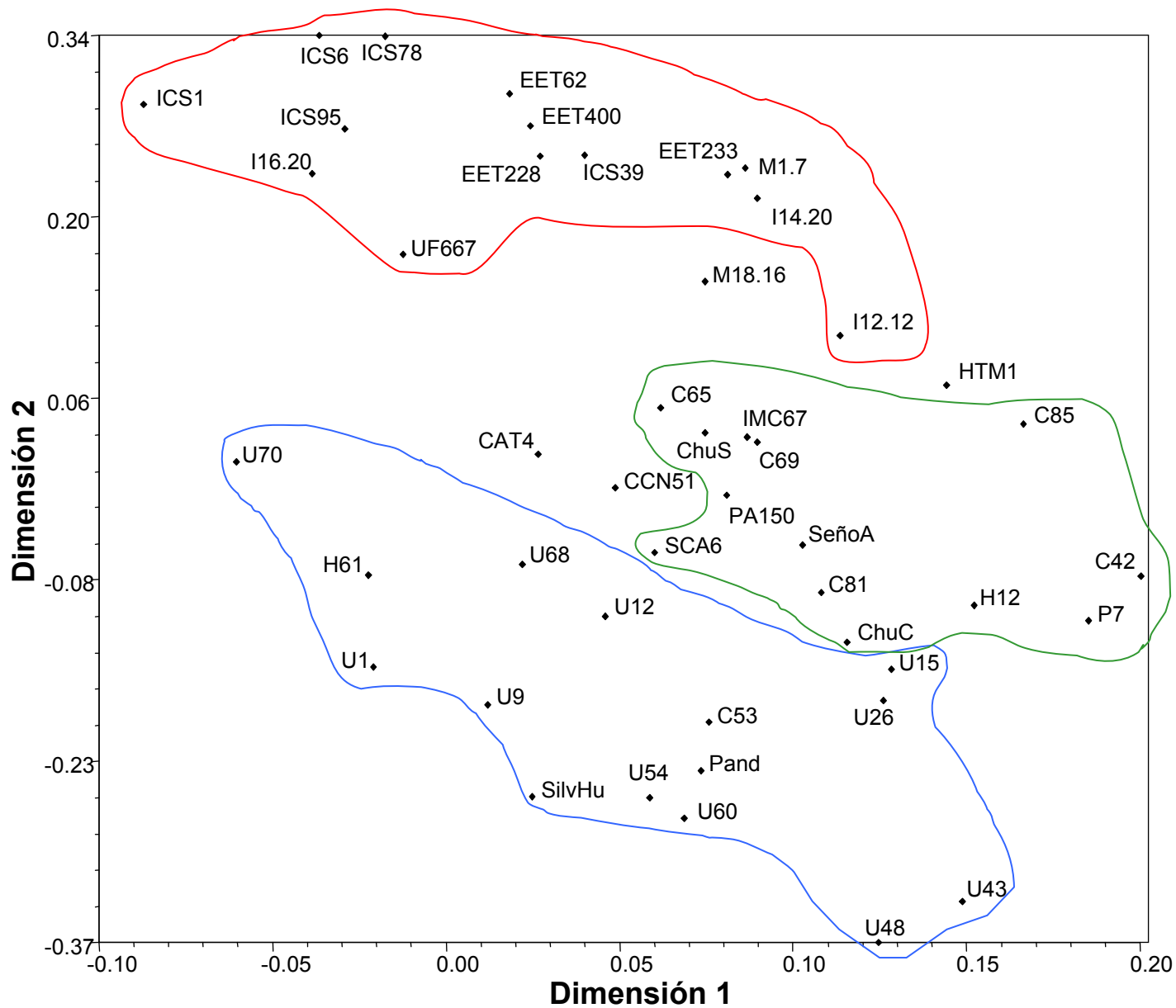
**Figura N° 12:** Fenograma de los híbridos y sus respectivos parentales generado con el coeficiente de asociación *Jaccard*. Los híbridos están resaltados con letra mayor. El coeficiente de correlación cofenética ( $r$ ) es de 0,77013.



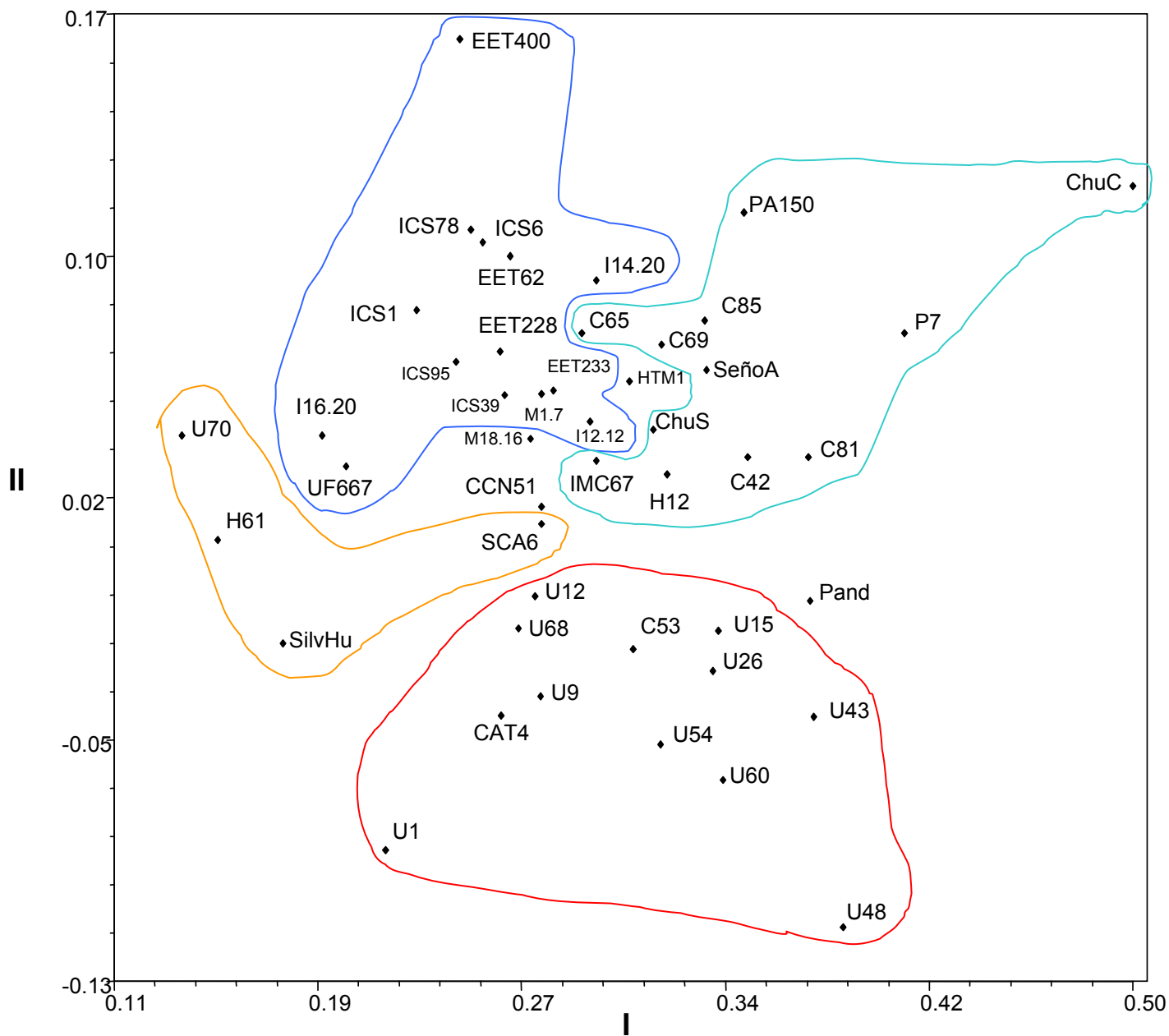
### 5.5.2 Ordenamiento

Se llevaron a cabo el Análisis de Coordenadas Principales y el Escalamiento Multidimensional no métrico para el ordenamiento de las OTUs en 2 dimensiones. En ambos casos, en el ordenamiento bidimensional, cuatro accesiones del total de 50 (UF29, H60, H37 y Seño) se ordenaron a una distancia muy alejada del resto. Entonces, para lograr una mayor resolución, se muestran sólo 46 de los 50 genotipos (Figuras N° 13 y N° 14). Sin embargo, los gráficos que contienen el total de genotipos se muestran en el Anexo N° 12 (Parte 1 y 2).

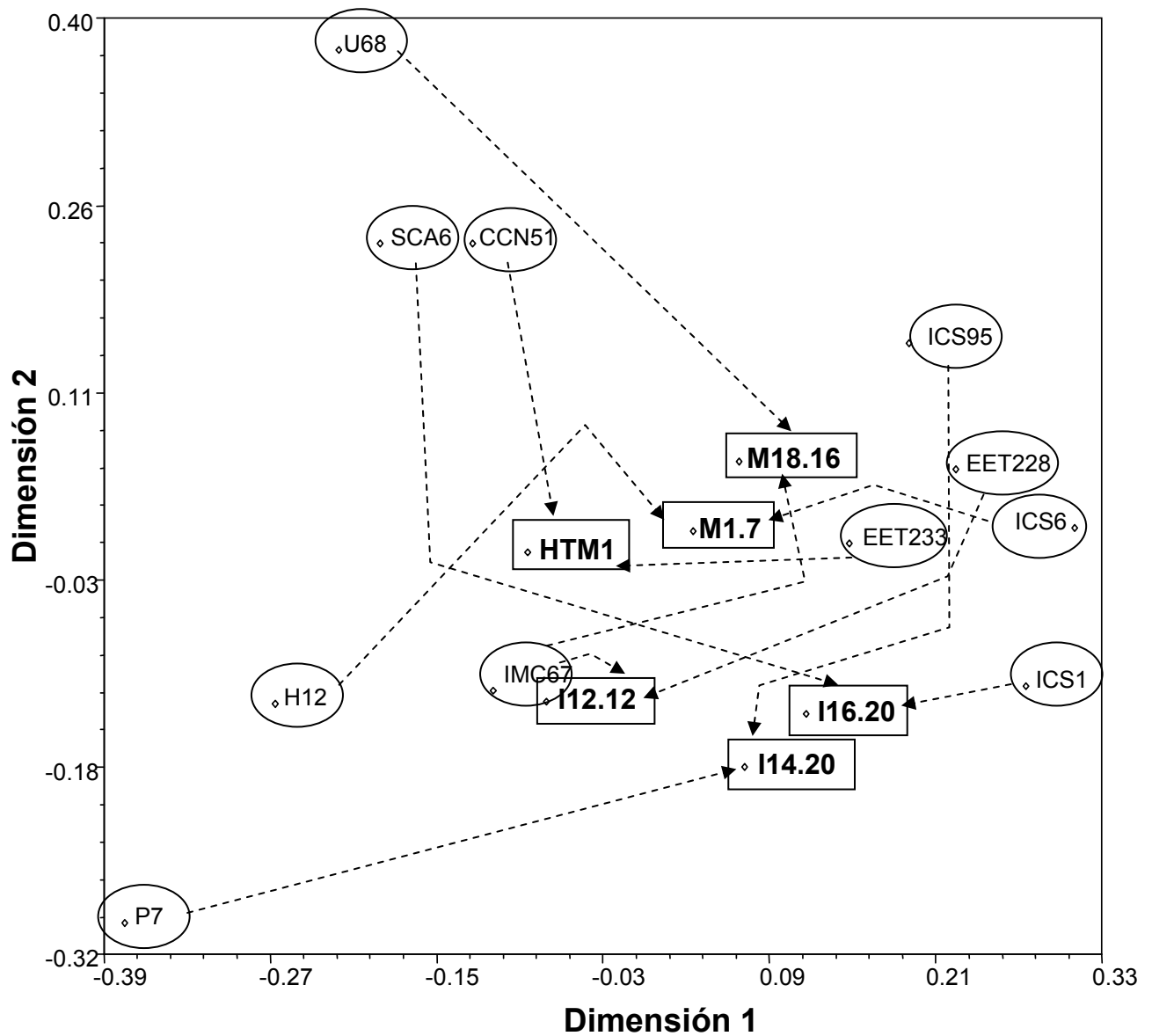
Se realizaron procedimientos similares de ordenamiento, esta vez con 6 Híbridos UNAS/Bioversity junto con sus respectivos parentales, en total 17 individuos. Para este caso, los gráficos de Análisis de Coordenadas Principales y Escalamiento Multidimensional no métrico se muestran en las Figuras N° 15 y 16, respectivamente. Se resaltaron los códigos de los híbridos con una mayor letra y encerrados en rectángulos. Los progenitores se encerraron en elipses o círculos, señalando con una flecha la relación con su progenie.



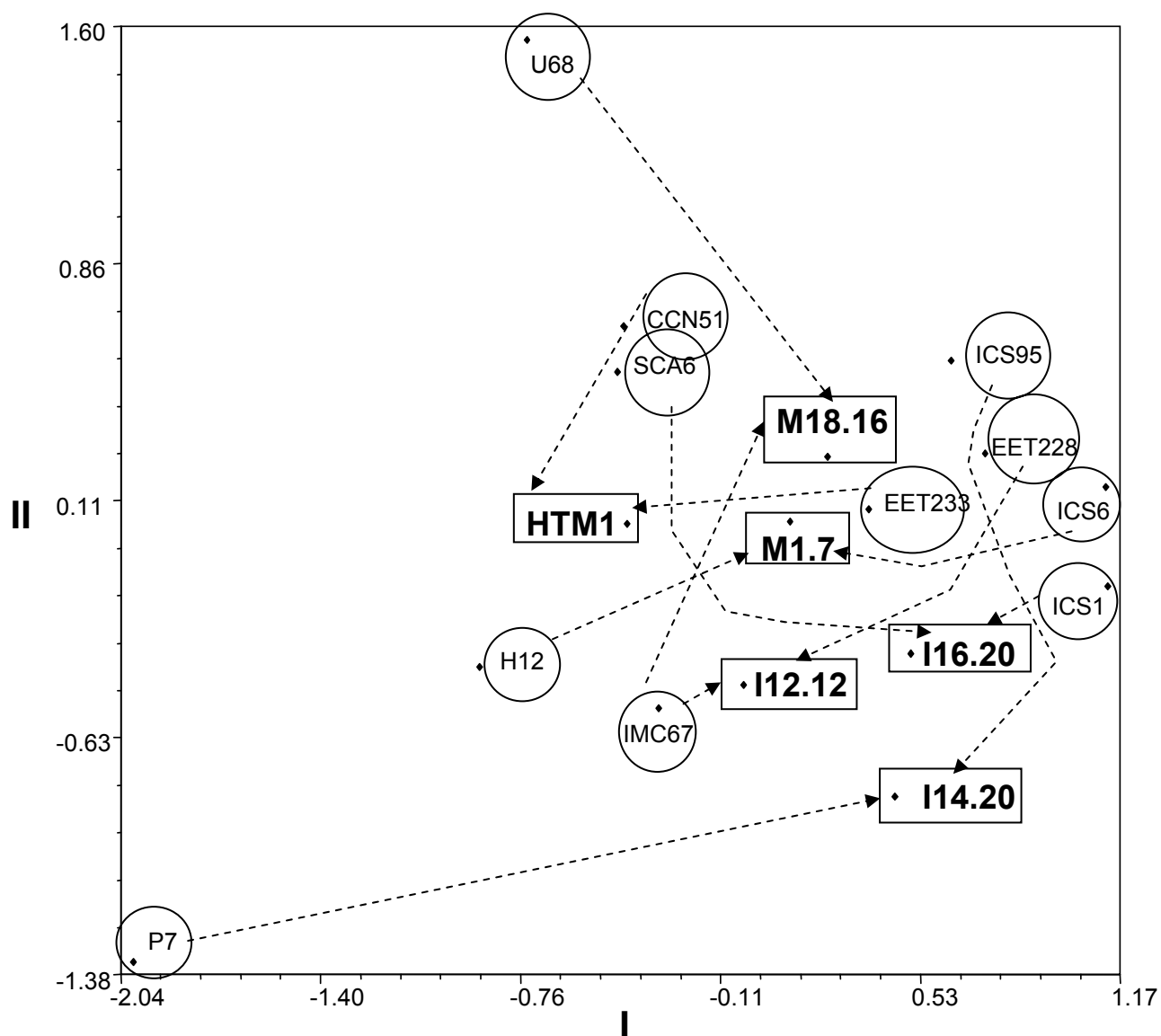
**Figura N° 13:** Análisis de Coordenadas Principales, coeficiente *Jaccard*. Con el fin de una mejor visualización, sólo se muestran 46 OTUs (se excluyen UF29, H60, H37 y Seño). Se han encerrado los genotipos más afines entre sí.



**Figura N° 14:** Escalamiento Multidimensional no métrico. Con el fin de una mejor visualización, sólo se muestran 46 OTUs (se excluyen UF29, H60, H37 y Seño). Se han encerrado los genotipos más afines entre sí. Valor de *Final STRESS2* = 0,10873.



**Figura N° 15:** Análisis de Coordenadas Principales de los Híbridos Bioversity/ UNAS (recuadros) y sus respectivos parentales (elipses), coeficiente *Jaccard*, se señala con una flecha la relación con su progenie.



**Figura N° 16:** Escalamiento Multidimensional no métrico de los Híbridos (recuadros) y sus respectivos parentales (círculos), coeficiente *Jaccard*, se señala con una flecha la relación con su progenie. Valor de *FinalSTRESS2* = 0,26212.

### 5.5.3 Estructura y variación Poblacional.

Se calcularon los índices descriptivos de polimorfismo para los grupos de accesiones: poblaciones y subpoblaciones. Estos se muestran en la Tabla N° 8.

**Tabla N° 8:** Cantidad de loci polimórficos y porcentaje correspondiente para cada población y subpoblación analizadas.

	Tamaño de la muestra	Número de loci polimórficos	% de loci polimórficos
Cusco	8	44	60,27
Huánuco	5	32	43,84
Junín	1	0	0
Loreto	4	13	17,81
San Martín	5	18	24,66
Ucayali	7	22	30,14
Brasil	1	0	0
Colombia	1	0	0
Ecuador	5	12	16,44
Trinidad y Tobago	5	8	10,96
Costa Rica	2	19	26,03
Subpoblaciones nacionales	6	60	82,19
Subpoblaciones internacionales	5	42	57,53
Poblaciones	2	66	90,41

Se calcularon los Índices de información Shannon y Diversidad génica de Nei con el programa POPGENE v1.32 (Yeh *et al.*, 1999) para cada población y subpoblación analizadas. Tales valores se ordenan y muestran en la Tabla N° 9.

**Tabla N° 9:** Índices de información Shannon y Diversidad génica de Nei para cada población y subpoblación analizadas.

<b>Entre individuos dentro de una subpoblación</b>				
	Índice de Información Shannon	desviación estándar	Diversidad génica de Nei	desviación estándar
Cusco	0,2793	0,2684	0,1814	0,1891
Huánuco	0,2611	0,3078	0,18	0,2167
Junin	0	0	0	0
Loreto	0,1022	0,2274	0,0697	0,158
San Martín	0,1414	0,2571	0,0965	0,1789
Ucayali	0,1619	0,2637	0,1091	0,1835
Brasil	0	0	0	0
Colombia	0	0	0	0
Ecuador	0,085	0,2039	0,0565	0,1408
Trinidad y Tobago	0,0634	0,1879	0,0435	0,1313
Costa Rica	0,1574	0,2672	0,1078	0,183

<b>Entre subpoblaciones</b>				
	Índice de Información Shannon	desviación estándar	Diversidad génica de Nei	desviación estándar
Clones nacionales	0,2995	0,2279	0,1853	0,1611
Clones internacionales	0,2123	0,2173	0,1276	0,1457

<b>Entre poblaciones</b>				
	Índice de Información Shannon	desviación estándar	Diversidad génica de Nei	desviación estándar
Nacionales vs. Internacionales	0,2977	0,2084	0,1799	0,1484

**Tabla N° 10:** Diferenciación genética poblacional respecto a todos los loci ( $G_{st}$ ).

<b>Comparación</b>	<b><math>G_{st}</math></b>
Dentro de las Nacionales	0,3798
Dentro de las Internacionales	0,7126
Entre todas las subpoblaciones	0,5502

De la misma manera, en el Anexo N° 13 se muestran los  $G_{st}$  de las comparaciones en pares de las subpoblaciones agrupadas *a priori*. Entre paréntesis se muestran el número de individuos por subpoblación.

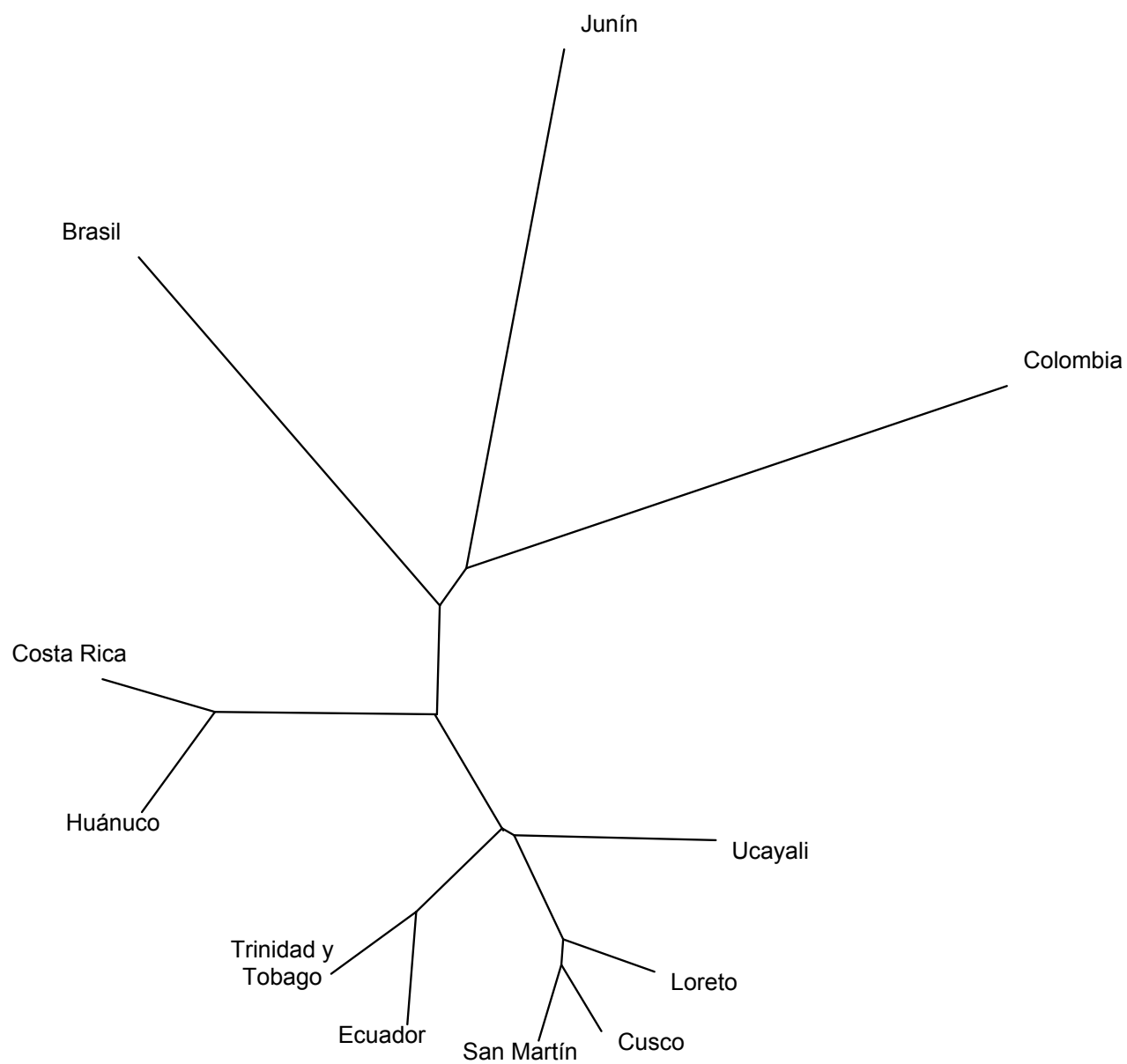
En la Tabla N° 11, se muestra la matriz de las Identidades Genéticas y Distancias Genéticas normalizadas (*unbiased*) de Nei. Las primeras se muestran encima de la diagonal, mientras que las Distancias Genéticas están debajo de la diagonal.

**Tabla N° 11:** Matriz de Identidades Genéticas y Distancias Genéticas normalizadas (*unbiased*) de Nei.

Subpoblación	Cusco	Huánuco	Junín	Loreto	San Martín	Ucayali	Brasil	Colombia	Ecuador	Trinidad y Tobago	Costa Rica
Cusco	*****	0,9489	0,8605	0,9691	0,9756	0,9279	0,8618	0,8659	0,9362	0,9158	0,9013
Huánuco	0,0524	*****	0,8255	0,9237	0,9134	0,8876	0,8310	0,8594	0,8972	0,8849	0,9611
Junin	0,1502	0,1918	*****	0,8743	0,8688	0,8564	0,8219	0,8082	0,8766	0,8630	0,8002
Loreto	0,0314	0,0794	0,1343	*****	0,9658	0,9146	0,8398	0,8743	0,9272	0,9133	0,8588
San Martin	0,0247	0,0905	0,1407	0,0348	*****	0,9570	0,8849	0,8825	0,9565	0,9378	0,8570
Ucayali	0,0749	0,1192	0,1550	0,0893	0,0440	*****	0,9127	0,8776	0,9368	0,9072	0,8460
Brasil	0,1487	0,1851	0,1961	0,1745	0,1223	0,0914	*****	0,7945	0,8848	0,8688	0,8150
Colombia	0,1440	0,1515	0,2129	0,1343	0,1250	0,1306	0,2300	*****	0,8483	0,8171	0,7496
Ecuador	0,0659	0,1085	0,1317	0,0756	0,0445	0,0653	0,1224	0,1645	*****	0,9658	0,8761
Trinidad y Tobago	0,0879	0,1223	0,1474	0,0906	0,0643	0,0974	0,1407	0,2020	0,0348	*****	0,9033
Costa Rica	0,1040	0,0397	0,2230	0,1522	0,1543	0,1673	0,2046	0,2882	0,1323	0,1017	*****

Con esta matriz generada con el POPGENE v1.32, se graficó un dendograma sin raíz con todas las subpoblaciones analizadas, el cual se muestra en la Figura N° 17. El gráfico se exportó y visualizado en el Treeview v1.6.0.





1

**Figura N° 17:** Dendrograma sin raíz de las subpoblaciones analizadas. Escala en la parte inferior izquierda. Se exportó el archivo del *software* POPGENE al Treeview para una mejor visualización.

El análisis molecular de variancia o AMOVA fue calculado con el software Arlequin v3.1 y se muestran los componentes de variación entre poblaciones (Nacional vs. Internacional), entre subpoblaciones y entre individuos dentro de subpoblaciones en la Tabla N° 12. La opción de AMOVA utilizada en este software fue calculado locus por locus. Los resultados son globales como promedios ponderados entre los loci analizados. El número de permutaciones ejecutada para el test de significancia fue de  $n = 1023$ .

**Tabla N° 12:** Análisis Molecular de Variancia con tres componentes: población, subpoblación y dentro de subpoblación.

Fuente de variación	Grados de Libertad	Suma de cuadrados	Componentes de variancia	Porcentaje de variación	P (significancia)
Entre Poblaciones	1	11,174	0,25978	6,46	$0,04399 \pm 0,00559$
Entre subpoblaciones dentro de cada población	9	50,719	0,67378	16,74	$0,00489 \pm 0,00203$
Dentro de la subpoblación	33	101,971	3,09004	76,80	0
TOTAL	44	163,864	4,02360		

Asimismo, con el mismo software se calcularon los estadísticos F promedio para todos los loci, estos parámetros se obtuvieron en función de las frecuencias alélicas:

$F_{SC}$  : 0,17295 (grado de diferenciación génica entre las subpoblaciones dentro de cada población).

$F_{ST}$  : 0,21463 (grado de diferenciación génica entre todas las subpoblaciones).

$F_{CT}$  : 0,05039 (grado de diferenciación génica entre poblaciones).

## 6. DISCUSIÓN

### 6.1 Elección de los genotipos

La muestra conformada por 50 individuos, involucró en su mayoría al grupo de Forastero Alto Amazónicos. El restante fue material Trinitario e híbridos de diferentes orígenes genéticos, algunos conocidos y otros desconocidos. No se dispuso de materiales de Tipo Criollo ni Nacional, aunque se incluyó en el análisis una accesión de Forastero Bajo Amazónico (Tabla N° 4).

El tamaño de la muestra es relativamente pequeña en comparación a recientes trabajos de caracterización molecular de cacao (Marita *et al.*, 2001; Parra *et al.*, 2001; Efombagn *et al.*, 2008; Motamayor *et al.*, 2008) que sobrepasan el centenar y a veces el millar de individuos. Sin embargo, Charters y Wilkinson 2000; Faleiro *et al.*, 2004; Zhang *et al.*, 2006; entre otros, también trabajaron con muestras relativamente pequeñas.

Los individuos a investigar se escogieron en base a dos criterios:

- Por las marcadas diferencias morfológicas entre accesiones.
- Por los caracteres de importancia que alberga cada uno de ellos.

Suponiendo con estos criterios que los individuos analizados serían genéticamente y molecularmente más divergentes y proporcionarían patrones de amplificación polimórficos.

La revisión de diversas tesis locales de caracterización morfológica (Bartra, 1993; López, 1993; Piñan, 1993; Gonzales, 1996; Vidal, 1997; García, 1998; Panduro, 1998; Rengifo, 2002; Guerrero, 2005; Aycachi, 2008) y otros trabajos internacionales (Engels, 1986; Arciniegas, 2005; Martinez, 2007) permitió recopilar información de caracteres importantes para el fitomejorador, por ejemplo, mayor rendimiento (representado por el índice de mazorca) y otros. Asimismo, la disponibilidad de los datos geográficos permitió hacer un análisis de algunos parámetros poblacionales.

## 6.2 Condiciones del ADN y la PCR

Zhang *et al.* (2006) reportan que la extracción del ADN en cacao es un proceso fundamental, ya que es uno de los factores que influyen en la reproducibilidad de los ensayos. En esta tesis, se siguió un protocolo para la extracción del ADN basado en el clásico trabajo de Doyle y Doyle (1987) obteniendo valores de pureza entre 1,20 a 2,39. El ADN considerado de alta pureza debe estar comprendido entre los valores de 1,80 y 2,00. Las concentraciones de ADN (ng/μl) obtenidas variaron desde 111 hasta 1802 ng/μl resultando en una amplia gama de concentraciones, pero de valores aceptables, dado que el PCR puede utilizar una cantidad muy pequeña de ADN por reacción. Se comprobó en electroforesis en agarosa (Figura N° 9) la integridad del ADN total extraído y se verificó que era necesario un tratamiento enzimático con ribonucleasa.

En relación a las condiciones del PCR, se utilizó un solo programa de amplificación para todos los iniciadores, que albergó una serie descendente de 7 temperaturas de alineamiento que permitieron acomodar a las diferentes temperaturas óptimas de la mayoría de los iniciadores ISSR.

De las pruebas de preselección con los 59 iniciadores, se escogió un grupo de los mejores iniciadores para realizar las reacciones definitivas de amplificación con las 50 accesiones. Dado que en las pruebas de preselección (con réplica biológica) no se observaron variaciones en los patrones de bandas, se siguió adelante con el proceso ya desarrollado para los demás individuos de cacao. Sarla *et al.*, (2003), en su análisis con ISSR de variedades de arroz, utilizaron la temperatura de 50° C para el alineamiento de seis diferentes iniciadores, y resolvieron las bandas con geles de agarosa al 1.8%.

De la lista de iniciadores con mejor amplificación (Tabla N° 6), siete de ellos son iniciadores anclados y los restantes no anclados. Los primeros tienen una secuencia que permite el alineamiento estable al extremo 3' o 5' de las secuencias microsatélites (De Vicente *et al.*, 2004); en cambio, a pesar de lo

que menciona Bornet y Branchard (2001), los iniciadores no anclados tienen posibilidad de unirse en diferentes lugares a lo largo de la secuencia complementaria al microsatélite. Por ello, para una mayor reproducibilidad se eligieron sólo iniciadores anclados (Tabla N° 7).

La electroforesis de poliacrilamida se realizó en cámaras artesanales, y frecuentemente se presentaron distorsiones o "sonrisas" a ambos lados del gel, por lo que se repitieron los ensayos con los individuos de estos carriles. Otro problema causado por la electroforesis fue la tinción con Hidróxido de sodio, la cual deja "background" en el gel, es decir, una coloración muy oscura que dificultó las lecturas de las bandas. Consecuentemente, la asignación de las bandas con los valores binarios (1,0) se realizó minuciosamente y las bandas dudosas se consideraron como datos perdidos, representados con el valor 2.

### **6.3 Valores descriptivos de los marcadores**

Fue necesario un tiempo considerable para dar el formato apropiado a los datos moleculares para su análisis con los paquetes Arlequin v3.1 y POPGENE v1.32. Debido a que son software libre o gratuito, tienen compatibilidad con un reducido número de programas de hojas de cálculo. En cambio, el NTSYS v2.1p (programa licenciado) si provee las facilidades para ingresar los datos binarios almacenados en un archivo MS Excel.

La confiabilidad de la asignación de datos y la reproducibilidad de las observaciones son cruciales para un sólido estudio de diversidad genética. Por ello, se vio en la necesidad del excluir loci con más del 10% de datos perdidos. A recomendación de Warburton y Crossa (2002), la cantidad de datos perdidos no debe sobrepasar el 15%, pero finalmente depende del criterio del investigador decidir tal cantidad.

Se seleccionaron un total de 73 bandas amplificadas, de las cuales 66 fueron polimórficas. Comparada con el trabajo de Charters y Wilkinson (2000), quienes lograron 45 bandas con 6 iniciadores ISSR para 82 accesiones de

cacao analizadas, la obtenida en esta tesis fue mayor. Sin embargo, en ese trabajo los autores no mencionan si realizaron una exclusión de loci con un porcentaje determinado de datos perdidos. Se podría asumir que no consideraron ningún loci con datos perdidos.

Sneath y Sokal (1973) describieron métodos para calcular el poder discriminante de un carácter o descriptor (de dos o multiestado) llamándolos coeficientes de separación. Este poder o valor discriminatorio permite al investigador tener una noción de la utilidad del carácter o los caracteres (loci) y de los iniciadores utilizados para diferenciar los *taxa*. Es posible medir la riqueza de los alelos para cada marcador, o la información que cada marcador imparte al estudio. Puede considerarse como la medida de la utilidad de cada marcador en distinguir un individuo de otro.

Se sumaron los PIC de todos los loci analizados para cada iniciador, proporcionando el Índice de Marcador para cada iniciador ISSR, el cual se basó en Guislain *et al.* (1999) (quienes lo reportan para RAPD) y se muestran en la Tabla N° 7 y en el Anexo N° 6. Los índices de iniciador más altos correspondieron a los iniciadores con códigos: UBC841, UBC842 y UBC884; cuyos índices fueron 2,43; 4,56 y 4,20 respectivamente, lo cual señala su valor potencial para un uso futuro en la caracterización molecular de cacao usando ISSR.

Se identificaron loci cuya frecuencia alélica de presencia ( $p$ ) fue menor o igual a 0,05 (Anexo N° 6). Estos alelos muy poco frecuentes entre la colección de los 50 individuos, fueron clasificados como “fragmentos raros” y cuantificados por *primer*, como se muestra en la Tabla N° 7. En forma complementaria se muestra en el Anexo N° 7 un ordenamiento en base a su origen de los genotipos que muestran fragmentos raros y la cantidad de éstos se señalan con una X por cada accesión. Se observa que se ha detectado una mayor cantidad en individuos de colectas nacionales, en especial de la colección Ucayali-Urubamba y en mucho menor medida en la Colección Internacional.

Conjuntamente con otros marcadores moleculares, estos resultados permitirían una identificación directa y reconocimiento de ciertos genotipos valiosos (Charters y Wilkinson, 2000). Marita *et al.* (2001), llevaron a cabo un estudio para determinar marcadores RAPD únicos asociados con genes relacionados a resistencia a enfermedades así como a otros caracteres de interés. Entonces, sería posible correlacionar a características de interés y a marcadores moleculares asociados a estos, con estudios más detallados y prolongados en segregación y herencia de los diversos genotipos de cacao para comprobar la estabilidad de su herencia en varias generaciones, y asegurar su utilidad como asistente molecular en un Programa de Mejoramiento.

#### **6.4 Agrupamiento**

Con el objetivo de definir las relaciones de similitud entre los 50 genotipos de cacao estudiados y encontrar entre ellos conjuntos consistentes y significativos de individuos, se realizó un análisis de agrupamiento utilizando el algoritmo UPGMA. Éste se deriva a partir de una matriz de distancias, calculada con un coeficiente de asociación (o llamado también de similitud). La literatura sobre taxonomía numérica (Sneath y Sokal, 1973; Crisci y López, 1983) reporta una numerosa cantidad de estos coeficientes. En esta tesis se utilizaron para los cálculos de la matriz de distancia, dos tipos: *Jaccard* y *Dice*, siendo éstos los más utilizados en estudios de diversidad molecular.

Se examinaron dos fenogramas obtenidos con tales coeficientes con la finalidad de lograr agrupaciones que expliquen de una manera más acertada las relaciones de similitud y las razones de su vinculación al correspondiente *cluster*. Entonces, al comparar los fenogramas generados con estos coeficientes (*Jaccard*: Figura N° 11; *Dice*: Anexo N° 10), se observó que la topología de ambos árboles era prácticamente idéntica. Sin embargo, los valores de similitud fueron mayores con *Dice* que con *Jaccard*. Esto significa que este último coeficiente muestra una mayor variabilidad entre los integrantes de la colección de cacao estudiada, dado que su composición

genética y geográfica es compleja. Charters y Wilkinson (2000) usaron el coeficiente *Jaccard* en una evaluación similar de una colección de cacao con marcadores ISSR. Entonces, con estos criterios, se eligió este coeficiente para todos los agrupamientos y ordenaciones y análisis poblacionales.

Del fenograma mostrado en el Figura N° 11, a un valor de 0,79 de similitud con el coeficiente *Jaccard*, se distinguieron 13 *clusters*, que se detallan a continuación:

**Cluster 1:** prácticamente la mayoría de las entradas internacionales e híbridos Bioversity/UNAS forman parte de este grupo, el genotipo discordante es el IMC67, que proviene de Loreto, Perú. Se notó que todos los miembros de este conglomerado comparten la característica de un número alto de semillas por mazorca, lo que animaría a pensar que esta cualidad pudo haber sido detectada molecularmente. En contraste, el PESE (peso de semilla) difiere considerablemente entre estas accesiones. El *cluster* se puede subdividir en 4 *subclusters*:

- **Subcluster 1.1:** formado por EET400. Es interesante cómo este único OTU se separa de los demás en el *cluster* principal. A pesar que comparte el código EET (Estación Experimental Tropical de Pichilingue, Ecuador) con 3 accesiones más de este estudio, pertenece a la raza Forastero alto amazónico, es decir, tiene un origen genético diferente.
- **Subcluster 1.2:** formado por EET228, EET233, M18.16, EET62, ICS6, ICS78, ICS95, ICS1. Los OTUs con códigos EET son de descendencia híbrida, al igual que las accesiones de códigos ICS. Éstos son Trinitarios, teóricamente son descendientes de (For. x Criollo). EET228 y el EET233 serían muy similares según lo determinado con ISSR. Genéticamente, el M18.16 sería diferente de los demás a pesar de ser de origen híbrido (IMC67 x U68, ambos Forasteros).
- **Subcluster 1.3:** constituido por ICS39, M1.7, I16.20, UF667, I12.12, IMC67, I14.20. En este sub-agrupamiento tenemos como integrantes a 3 híbridos del proyecto Bioversity/UNAS, los cuales tienen parentales ICS, por lo cual es de esperar que estos genotipos se agrupen. El genotipo que discordaría de los demás en el grupo es el IMC67, el cual es de raza Forastero Alto amazónico y tiene un origen geográfico diferente de los



demás. Morfológica y agronómicamente (según los datos recopilados en el Anexo N° 1 y N° 2), los genotipos ICS39, UF667 e IMC67 son muy distintos entre sí; por ejemplo, en la característica índice de mazorca estos genotipos tienen: 13, 15.30 y 16.36, respectivamente.

- **Subcluster 1.4:** formado por ChuS. Singularmente este genotipo proveniente de Satipo (Junín) está formando parte del *cluster* 1; sin embargo, se separa del resto de genotipos, de los cuales tiene diferencias geográficas. No se han determinado sus características morfológicas ni agronómicas por lo que proviene de una colecta reciente.

**Cluster 2:** formado por C42, HTM1, C65, PA150. Este *cluster* está formado por genotipos muy divergentes, un híbrido HTM1 (cruza de internacionales CCN51 x EET233) de origen muy complejo, C42 y C65 también provienen de un origen híbrido (For. x Tr.) colectados en parcelas de agricultores de la región San Martín. Como miembro contrastante se encuentra el PA150, el cual es Forastero y divergiría de los demás. Si analizamos la genealogía del HTM1, su padre CCN51 deriva del IMC67, ICS95 y una variedad canelo, esto puede sugerir que la agrupación de HTM1 con variedades peruanas no estaría fuera de lugar.

**Cluster 3:** formado por U1, U9, U12, U15, CAT4. Este *cluster* sería uno de mayor coherencia, ya que agrupa genotipos U# (Ucayali) colectados en zonas muy cercanas, probablemente compartan un origen genético común detectado por los marcadores. El genotipo CAT4 (Catongo) proveniente de Brasil, es de raza Forastero Bajo amazónico. Posee cotiledones de color blanco, una de las características principales de los Criollos. Entonces, se puede inferir que esta agrupación comparte otros elementos moleculares que se han detectado a nivel de los ISSR. La cercana similitud entre CAT-4 con los clones Ucayali podría deberse a que los genotipos Ucayali en un pasado hayan sido trasladados por agricultores vía fluvial de sus lugares hasta el río Amazonas y a partir de aquí cruzaron territorio brasileño y posteriormente seleccionado en la Amazonía baja. Esta es una hipótesis que quedaría comprobarla mediante estudios de flujo génico.

**Cluster 4:** formado por U26, U54, SCA6, CCN51, C53, U68. Este grupo contiene genotipos muy diversos geográfica y genéticamente. El U26, U54 y U68, son Forastero Alto Amazónicos; sin embargo, los dos últimos fueron colectados en la cuenca del río Urubamba y el primero en la cuenca del río Ucayali. El C53 es de origen híbrido (For. x Trin.), pero se localiza en las localidades cercanas a la cuenca de Ucayali, lo que nos sugiere que el parental Forastero de este híbrido, sea el componente común con los genotipos ya mencionados. SCA6 es un genotipo muy utilizado en los programas de mejoramiento genético, debido a su aporte a la resistencia a “escoba de bruja”. El híbrido CCN51 posee esta misma característica y descende del IMC67, que es un genotipo colectado geográficamente cercano al SCA6.

**Cluster 5:** formado por U43, U48. Ésta agrupación algún tendría fundamento, ya que los dos genotipos fueron colectados en la cuenca del río Ucayali, geográficamente cercanos y supondrían un origen genético común. Aunque en las colectas lo que se pretende es colectar en una región la mayor cantidad de diversidad posible, debido al flujo génico, los individuos pueden compartir características comunes.

**Cluster 6:** formado por SilvHu, SeñorA, H12, Pand. El SilvHu (árbol de cacao silvestre del Huallaga) podría tener similitudes genéticas con H12 (Colección Huallaga); sin embargo, a pesar de estar juntos, SeñorA (Señorita Achoccha) y Pand (Pandora) son diferentes geográficamente con el resto y entre sí. El primero es conocido como cacao “chuncho” del Cusco y el segundo es un cacao de Colombia con un regular Índice de mazorca (valor de 16,9).

**Cluster 7:** formado por U60, C69, C81. Agrupación entre genotipos morfológicamente divergentes: U60 es Forastero, C69 y C81 son de descendencia híbrida For. x Tr. Estos últimos tienen diferente color de fruto inmaduro: rojo y verde, respectivamente, lo que los hace diferentes en términos de raza. Otras características morfológicas entre estas accesiones dan que pensar que la relación de similaridad entre ellas no se da a un nivel fenotípico.

**Cluster 8:** formado por H61, constituyendo la colección Huallaga. Los miembros de esta colección provienen de colectas en zonas muy cercanas a la localidad de Leoncio Prado, es decir de un área limitada. Según Vidal (1997), la diversidad genética que poseen es morfológicamente baja; teoría reforzada con los resultados de Zhang *et al.*, (2006), quienes trabajaron con SSR. Sin embargo, utilizando los ISSR, los 4 genotipos Huallaga analizados en este estudio divergen molecularmente, como se revela en el fenograma.

**Cluster 9:** formado por C85, P7, ChuC. El C85 es un cacao tipo “chuncho” el cual se agrupó con su similar del Cusco: ChuC. De estas accesiones, aún no se disponen de datos morfológicos y agronómicos; aunque comparten orígenes geográficos y mediante los ISSR se detectaron similares. El genotipo P7, Pound 7 de Iquitos, es un Forastero Alto amazónico que de alguna manera divergiría genéticamente de los anteriores, sin embargo, se agrupa con ellos. El chuncho ChuC posee una mazorca muy diferente al de Señora, que está agrupado en el *cluster* 6.

**Cluster 10:** formado por U70. Proveniente de la cuenca del río Urubamba, este cacao Forastero Alto Amazónico posee con cotiledones blancos (García, 1998), lo cual lo hace muy interesante. Esto evidenciaría la hipótesis de genes mutantes cuya expresión hereditaria se traduce en el color de los cotiledones. Se podrían conseguir más de estos genotipos interesantes en la cuenca del Río Urubamba o sus alrededores. La presencia y ubicación de los genotipos U70 y CAT4 en el fenograma (cacao Forasteros con cotiledones blancos), puede ayudar a reforzar la hipótesis de que el cacao es originario de los bosques tropicales de América del sur, y que los cacao “Criollos” de Centroamérica han surgido de la domesticación de ciertos individuos traídos desde el sur.

**Cluster 11:** formado por UF29, H60. A pesar de sus orígenes híbridos con Forastero, el primero posee un origen materno Nacional y el segundo Trinitario. Lamentablemente, no se cuenta con mucha información de H60; asimismo, y como resultado de una exploración poco profunda de la

diversidad genética molecular con unos pocos iniciadores, los agrupamientos probablemente requieran mayor resolución.

**Cluster 12:** formado por H37 de la colección Huallaga. Tiene un origen híbrido For. x Tr. y comparte muchas características morfológicas entre los miembros de la colección Huallaga. Su divergencia del resto de los miembros de esta colección sugiere que posee variación a nivel molecular.

**Cluster 13:** formado únicamente por Señorita. Conocido como cacao “señorita” pertenece al grupo For. Alto Amazónico y localmente llamado “chuncho”. No se cuenta con mucha información morfológica para este individuo, ya que proviene de una colecta reciente. Sin embargo, se puede presumir por su ubicación en el árbol fenético que posee una amplia diversidad molecular. No se agrupa con los demás individuos “chunchos”. Señorita y U70 (del *Cluster* 10), son ambas accesiones provenientes de poblaciones de cacao “chuncho”; sin embargo, se muestran molecularmente lejanos. En un estudio reciente sobre la diversidad de cacao (M & O Consulting, 2008), se describe que existen marcadas diferencias morfológicas entre muchas poblaciones de cacao chuncho en el Cusco. Esta heterogeneidad podría deberse a costumbres ancestrales de los indígenas de colectar mazorcas de varios árboles silvestres y sembrar sus semillas híbridas.

En el caso del análisis de los híbridos y sus parentales (Figura N° 12), se puede observar que no se logró agrupar los híbridos y sus respectivos padres. Esto podría deberse al hecho de que la recombinación génica juega un papel preponderante en la mezcla de los alelos y diferentes secuencias génicas al realizar la cruce entre los dos genotipos parentales. No obstante, se deben ensayar más iniciadores y diseñar un experimento de segregación de los marcadores moleculares para establecer concretamente tales relaciones. Charters y Wilkinson (2000) utilizando 9 *primers* lograron agrupar parentales y su progenie en sus respectivas ramas, dos grupos derivados de la autofecundación de una variedad altamente homocigota y de una variedad altamente heterocigota. La primera formaba con su progenie un *cluster* con

similitud al 100%, y la última formaba un *cluster* más variable que se expande a partir de un 74% de similitud.

Lachenaud (1997), comentó que muchos de los trabajos realizados para categorizar o clasificar el cacao tuvieron las siguientes "debilidades": la suposición de que ciertos clones (individuos o accesiones) pertenecen a particularmente uno de los llamados "Grupos Básicos": Forastero, Criollo o Trinitario (propuestos por Cheesman, 1944). Asimismo, el sesgo generado por el análisis estadístico jerárquico de datos no balanceados (números diferentes de muestra). Es por ello, que se pueden encontrar contradicciones en clasificaciones o estudios evolutivos del cacao.

Parra *et al.* (2001) y Parra *et al.* (2002) evidenciaron la variabilidad existente entre estos tipos de cacao como consecuencia de la hibridación natural que ha ocurrido en la zona. Encontraron poca asociatividad entre agrupamientos y localización geográfica. Asimismo, Faleiro *et al.* (2004) encontraron que el análisis de agrupamiento dio lugar a *clusters* con materiales de diferentes orígenes amazónicos, sin ninguna tendencia a la regionalización de la variabilidad genética. Guerrero (2005) realizó un estudio de taxonomía numérica con caracteres cualitativos y cuantitativos del cacao, con la finalidad de determinar las relaciones entre los miembros de la colección Ucayali – Urubamba y observó que existe una muy baja correlación entre el agrupamiento por caracteres cualitativos comparado con el agrupamiento realizado con caracteres cuantitativos.

En el estudio molecular de una muestra de 952 accesiones de cacao amazónico, Motamayor *et al.* (2008) no encontraron apoyo para las hipótesis propuestas por otros autores de diversificación ribereña o de centros de refugio de especies amazónicas, dado que los individuos agrupados en el mismo *cluster* pueden ser encontrados en diversos ríos o cuencas principales. La distribución geográfica de los *clusters* no corresponde a los hipotéticos centros de refugio propuestas para otras especies en la región.

Motamayor *et al.* (2008) proponen que el patrón de diferenciación de las poblaciones estudiadas de cacao podría estar ligado a potenciales barreras de dispersión creadas por cordilleras "primitivas" también llamadas "paleoarcos". Los individuos de varios *clusters* se pueden ubicar en una misma localidad sin ser divididos por alguna potencial barrera (río o cordillera). Es el caso de Iquitos, donde varios ríos principales usados para transporte convergen, la presencia de poblaciones distintas de cacao podría deberse a una antigua o actual intervención humana. Aunque no se han realizado investigaciones al respecto, se puede plantear la hipótesis que el flujo génico entre diferentes zonas de la Selva Amazónica existe gracias al hombre, dado que el fruto de cacao es indeshiscente y la dispersión por polinizadores muy limitada (Días, 2001). Esto explicaría parcialmente el aporte y la similitud molecular entre algunos individuos de cacao cuyo origen geográfico no es el mismo.

## 6.5 Ordenamiento

El ordenamiento es un método de análisis multivariado que complementa al agrupamiento. De esta manera se obtienen representaciones gráficas que ayudan a interpretar de manera intuitiva las relaciones de los diferentes grupos de OTUs. Se utiliza como una herramienta exploratoria para probar una hipótesis. Se pueden contrastar con conocimientos disponibles de las muestras en cuestión, y en la medida de lo posible, con información adicional relacionada con la pregunta biológica objeto de la investigación (De Vicente *et al.*, 2004).

Se decidió realizar dos tipos de ordenamiento: Análisis de Coordenadas Principales y Escalamiento Multidimensional no métrico. Ambos sistemas se han utilizado para ordenar las OTUs en dos o tres dimensiones en diferentes trabajos de caracterización y diversidad molecular de cacao (Marita *et al.*, 2001; Faleiro *et al.*, 2004; Dos Santos *et al.*, 2005; Zhang *et al.*, 2006). De todos estos trabajos revisados, el método más utilizado fue de Escalamiento Multidimensional. Éste logra optimizar más eficientemente los resultados

cuando hay presencia de datos perdidos, a diferencia del Análisis de Coordenadas Principales (Rohlf, 2000).

En la comparación de los gráficos bidimensionales generados y que muestran las 50 accesiones en total (Anexo N° 12, parte 1 y 2) se pueden distinguir que cuatro de ellas: UF29, H60, H37 y Seño, son ubicadas a una gran distancia del resto de OTUs. Por ello, para visualizar de una manera más resuelta la disposición de los individuos, se graficaron sólo 46 accesiones con escalas X-Y más apropiadas (Figura N° 13 y 14). En estas Figuras, se puede observar que ambos métodos de ordenamiento lograron discernir de manera poco exhaustiva entre individuos de la colección internacional y la colección nacional, a excepción de CCN51, CAT4, Pandora y UF29 (éste último se muestra en el Anexo N° 12). De manera similar que en el agrupamiento UPGMA, la mayoría de híbridos Bioversity/UNAS se ubican en las inmediaciones de la zona de la colección internacional, demostrando que son descendientes de éstos y comparten algún grado de similitud molecular.

Particularmente, se puede notar que en ambos: agrupamiento y ordenamiento, el CCN51 se ubica muy cerca del SCA6, ambos árboles poseen tolerancia a la enfermedad “escoba de bruja” causada por *Crinipellis perniciosus*. Sin embargo, el primero es además susceptible a *Phytophthora palmivora* y *Moniliophthora roreri* (García, 2000). Los genotipos U15 y U26 se agrupan muy cerca en los dos métodos de ordenamiento, así como también en el Fenograma. Estas accesiones comparten un origen geográfico común y ambas poseen una importante característica que es un alto contenido de grasa en la semilla, sugiriendo su potencialidad como parentales en programas de mejoramiento genético.

En ambos métodos de ordenamiento bidimensional con los híbridos (Figura N° 15 y 16), las ubicaciones de los híbridos y sus respectivos parentales fue espacialmente muy similar. En el caso de I12.12 y I16.20 se ubican cerca de sus respectivos padres, pero alejados de sus respectivas madres. Los genotipos I14.20, M1.7, M18.16 se ubican lejos de sus parentales. En

cambio, HTM1 se ubica relativamente lejos de sus parentales. En todos los casos, los genotipos parentales se han ubicado en posiciones alejadas unos respecto a otros, lo que se merece reconocer ya que de esta forma se obtiene una descendencia híbrida con una mayor ganancia genética (Dos Santos *et al.*, 2005).

## 6.6 Validación interna

Con el fin de demostrar la solidez de los gráficos de agrupamiento generados, se utilizó la prueba conocida como la correlación de Mantel de dos vías, en el software NTSYSpc v2.1p, mediante la cual se puede demostrar, comparando la matriz de similitud versus la matriz cofenética, que hay una adecuada representación de las similitudes genéticas por parte del fenograma (Crisci y López, 1983; Rohlf, 2000; De Vicente *et al.*, 2004). Entonces, se generaron los gráficos que se muestran en el Anexo N° 11, los cuales señalan la tendencia a una relación de linealidad entre las matrices de similitud y cofenética. En los fenogramas generados con los 50 accesiones (Figuras N° 11 y Anexo N° 10), los coeficientes de correlación cofenética o  $r$ , resultaron siendo mayores a 0,9. Esto indica que la técnica de agrupamiento UPGMA origina una distorsión muy baja, o dicho en otras palabras, que los agrupamientos en los fenogramas representan de manera fiel los valores de similitudes genéticas entre las 50 accesiones de cacao.

En el caso del fenograma de la Figura N° 12, en el que se analizan sólo 17 individuos, se obtuvo un valor  $r$  igual a 0,77013, que revela que este agrupamiento representa pobremente a la matriz de similitud de tales accesiones. En operaciones no mostradas en la tesis, se ensayaron otros algoritmos de agrupamiento como el de ligamiento completo y simple, obteniendo valores de  $r$  iguales a 0,68814 y 0,71932, respectivamente. Éstos son evidentemente menores al obtenido con el UPGMA, por tanto, no son convenientes para representar estas relaciones de similitud. Osorio *et al.* (2003) a través de un estudio de diversidad genética en cacao, compararon tres algoritmos de Agrupamiento: Ligamiento Completo, Ligamiento



Promedio y Ligamiento Simple, y si bien no obtuvieron entre ellos diferencias significativas entre los coeficientes de correlación, estos valores fueron mayores con el Ligamiento promedio para todos los casos examinados.

Con respecto al Escalamiento Multidimensional no métrico, se evaluó la precisión del ordenamiento de las OTUs con el valor de *FinalSTRESS2* (Rohlf, 2000). Este parámetro es seleccionado por defecto en el mencionado análisis. En el ordenamiento de las 50 accesiones, el valor de *FinalSTRESS2* se obtuvo un valor de 0,10873, que es calificado como bueno, según la escala presentada en la sección 3.11 de la Metodología. En cambio, en el ordenamiento del grupo de híbridos UNAS/Bioversity y sus parentales, se obtuvo el valor de 0,26212, que puede ser considerado como aceptable.

## 6.7 Análisis Poblacional

En las estimaciones de todos los parámetros poblacionales, se debe observar que debido al pequeño tamaño de la muestra, las subpoblaciones Junín, Brasil y Colombia cuentan con un sólo individuo, en consecuencia, cabría esperar desviaciones en las estimaciones. Para realizar los análisis poblacionales, se realizó la suposición previa de que las poblaciones de cacao están en Equilibrio Hardy Weinberg; ésta presunción estaría apoyada en parte, ya que el cacao es una especie principalmente alógama y gracias a esto son poblaciones muy heterocigotas. Sin embargo, se presentará distorsión para este equilibrio, dado que existe presión de selección, deriva génica y migración (flujo génico) en las poblaciones de cacao.

El análisis de los datos generados por la amplificación ISSR está limitado por la naturaleza dominante de sus bandas. Se pueden determinar algunos parámetros genéticos y poblacionales con el riesgo de cometer un gran sesgo. Por ello, no se tomaron en cuenta los valores de **A** (número de alelos por locus), **A<sub>p</sub>** (número de alelos por locus polimórfico), **H<sub>e</sub>** (proporción de heterocigosis esperada) y **H<sub>o</sub>** (proporción de heterocigosis observada). Estos parámetros genéticos descriptivos se calculan en base a las ausencias de bandas, lo que no es recomendable debido a que se asume que tales

caracteres comparten una idéntica mutación ancestral, o en otras palabras, que fuesen caracteres homólogos, causando distorsiones en la interpretación genética.

#### **6.7.1 Porcentaje de loci polimórficos**

Se calculó el porcentaje de loci polimórficos directamente de los datos (conteo de presencias y de ausencias) para subpoblaciones y poblaciones (Tabla N° 8). Observando estos resultados, en la colección Nacional el polimorfismo llega al 82,19 % mayor que el 57,53 % de la colección Internacional. Entre las subpoblaciones, la región donde encontramos un mayor polimorfismo, es el Cusco con 60,27 %, seguido de Ucayali con 30,14 %. Gracias a estos valores se puede dar una idea superficial que los mayores niveles de diversidad en términos de polimorfismo se encuentren en tales regiones. Es necesario continuar con los estudios con muestras más grandes. En el caso de las subpoblaciones Junín, Brasil y Colombia, que poseen un único individuo, no fue posible la determinación de sus % de loci polimórficos.

#### **6.7.2 Índices de Shannon y Diversidad Genética de Nei**

Realizando una inspección de la Tabla N° 9, se puede distinguir que los índices de Shannon varían desde 0,0634 a 0,2793 y los índices de Nei van desde 0,0435 a 0,1814. Tres subpoblaciones no computaron estos parámetros: Junín, Brasil y Colombia, por estar representadas por un solo individuo. Los valores más altos de diversidad genética son la de Cusco y Huánuco, como lo indican ambos parámetros. Esto sugeriría al fitomejorador que hay un potencial nivel de diversidad genética y molecular en tales zonas geográficas y apoyaría en la proyección de diferentes estrategias para aprovechar el germoplasma que albergan. En el caso de las subpoblaciones analizadas, los valores de Nei no son numéricamente como los Índices de Shannon; sin embargo, se correlacionan positivamente en todos los grupos. Adicionalmente, se encontró una mayor diversidad genética entre las subpoblaciones dentro de las dos poblaciones analizadas (Nacional e

Internacional) con valores Shannon de 0,2995 y 0,2123, respectivamente. Cabe añadir que los valores de desviación estándar son relativamente altos. Sounigo *et al.*, 2005, quienes realizaron una caracterización dual con isoenzimas y RAPD de individuos del Banco de Germoplasma de Trinidad y Tobago, reportaron índices de Shannon que variaron desde 0,09 hasta 0,42 con desviaciones estándar entre 0,17 y 0,29 obtenidos con RAPD. Asimismo, estos autores comparan y discuten estos valores con los obtenidos por Isoenzimas y determinaron que poseen una correlación altamente significativa.

### **6.7.3 Coeficiente de diferenciación genética $G_{st}$**

Los valores de  $G_{st}$  indican la proporción de la diversidad genética entre las poblaciones comparadas y permitiría ser la guía para definir la cantidad de individuos que una muestra debe contener. A una mayor diversidad genética, mayor número de individuos (Sounigo *et al.*, 2005). Esto también permitiría dirigir las estrategias de exploración para coleccionar y enriquecer la diversidad genética en un Banco de Germoplasma.

En el Anexo N° 13, se muestra la comparación uno a uno de los  $G_{st}$  entre subpoblaciones. Cusco se diferencia en forma moderada de las subpoblaciones Internacionales, excepto de Brasil. Huánuco, se diferencia considerablemente de Ecuador y Trinidad, y medianamente de Loreto, San Martín y Ucayali. En su trabajo con microsatélites, Zhang *et al.* (2006), confirmaron una diferenciación genética entre la colección Ucayali-Urubamba, compuesta de individuos provenientes de Ucayali y Cusco, respectivamente. Las comparaciones entre Junín, Brasil y Colombia subpoblaciones con otras no se consideran para discusión para evitar la determinación de valores sesgados.

En el trabajo de Sounigo *et al.* (2005), los valores de  $G_{st}$  obtenidos para 23 poblaciones utilizando RAPD fue de 0,35; el cual es menor que el obtenido en este trabajo: 0,55 (Tabla N° 10). Esto podría sugerir una sobreestimación de este parámetro, ya que se utilizaron subpoblaciones con pocas muestras

y en algunos casos con un solo individuo. Aunque estos autores reportan que obtuvieron un valor menor comparado al reportado previamente por Russell *et al.* (1993) (citado por Sounigo *et al.*, 2005), quienes obtuvieron 0,42 en un estudio de tres poblaciones de cacao.

#### 6.7.4 AMOVA

La diferenciación genética entre poblaciones fue baja, con un 6,46 %, significando que este pequeño porcentaje de variación genética residiría entre las colecciones Nacional e Internacional. Asimismo, entre las subpoblaciones dentro de cada población, el porcentaje de variación fue de 16,74 %, y estos dos porcentajes de variación fueron menores en comparación con el nivel de variación entre individuos de cada subpoblación: 78,54 %. Los porcentajes de variación fueron altamente significativos ( $p < 0,01$ ) en los dos últimos casos y significativo ( $p < 0,05$ ) en el primer caso (Tabla N° 12). Estos resultados podrían concordar con los de Motamayor *et al.* (2008), quienes determinaron que el mayor porcentaje de variación (44,6 %) residía entre los individuos de los *subclusters* (es decir, dentro de las subpoblaciones). Mientras que entre los *subclusters* dentro de los *clusters* (es decir, entre las subpoblaciones) hubo un 17,3 %. Por último, entre los *clusters* (es decir, entre poblaciones) obtuvieron un valor de 38,1%. Ellos aclaran que este último porcentaje es alto comparado con otros trabajos similares, atribuyéndolo a la exclusión para el análisis de genotipos atípicos.

El AMOVA particionó la variación observada en componentes dentro y entre poblaciones usando distancias genéticas. El componente de variancia entre poblaciones es llamada  $\Phi_{st}$ , y es un análogo de  $F_{st}$  y  $\theta$ . Este método no fue originalmente diseñado para analizar datos puramente dominantes (como los RAPD, AFLP o ISSR), y realmente se deben considerar los datos como haplotipos moleculares. Motamayor y Lanaud (2002), mencionan que las poblaciones y cultivares de cacao en Sudamérica poseen una diversidad intrapoblacional mayor que la interpoblacional. En contraste, se ha encontrado que en México (donde se atribuye la región de domesticación del

cacao), sucede lo inverso, es decir que la diversidad entre dos poblaciones es mayor que la encontrada dentro de cada región.

El  $F_{ST}$  fue estimado en 0,23202 y según De Vicente *et al.* (2004), indicaría que las poblaciones analizadas difieren considerablemente, ya que está comprendido en el rango de 0,15 a 0,25. Asimismo, se puede interpretar indirectamente que a un mayor  $F_{ST}$  el flujo génico entre las subpoblaciones es menor.

## 7. CONCLUSIONES

- 1 Con el uso de los marcadores moleculares ISSR se logró determinar el nivel de diversidad molecular existente en la colección estudiada de 50 individuos de cacao, utilizando el coeficiente de similitud Jaccard.
- 2 Se logró encontrar que en el *cluster* 1 las accesiones comparten característica de alto NUSE, sin embargo, para el resto de accesiones y características no se halló una estricta relación entre la información morfológica, geográfica y molecular, manifestada en los fenogramas y representaciones bidimensionales. Por otro lado, no se logró asociar los híbridos Bioversity/UNAS y sus respectivos parentales con el análisis de agrupamiento y ordenación generados a partir de los datos moleculares ISSR.
- 3 Indicados por los parámetros de Shannon y de Nei, se observaron valores muy considerables de diversidad genética en las regiones Cusco y Huánuco, en comparación al resto de subpoblaciones.
- 4 Se halló a través del AMOVA, una gran variación molecular a nivel intrasubpoblacional (elegidos *a priori* según su origen geográfico), concordando con trabajos previos de diversidad molecular.

## **8. RECOMENDACIONES**

- 1 Ensayar los 5 iniciadores ISSR probados, y utilizar los 8 restantes (Tabla N° 6) en una colección nacional de germoplasma mayor.
- 2 Las representaciones fenéticas (por agrupamiento y ordenamiento) de los 50 individuos de cacao se pueden considerar como un criterio más para dirigir los cruzamientos en los programas de mejoramiento genético.
- 3 Aislar y estudiar las secuencias ISSR relacionadas al grupo del cluster 1, en el cual todos sus miembros comparten la característica de un alto número de semillas (NUSE).
- 4 Realizar estudios de segregación de los marcadores ISSR entre parentales y descendientes, para reforzar las hipótesis de herencia de esos caracteres.

## 9. REFERENCIAS BIBLIOGRÁFICAS

- ADRIAZOLA, J.L. 2003. Producción del Alimento de los Dioses *Theobroma cacao* L. Monografía. Universidad Nacional Agraria de la Selva. Tingo María. Perú. 81 pp.
- ARCA, M. 2000. Capítulo I: La producción del cacao en el Perú. En: Arca B., M. (Ed.) El Cultivo del Cacao en la Amazonía Peruana. Presidencia de la República. Ministerio de Agricultura, Lima - Perú.
- ARCINIEGAS, A.M. 2005. Caracterización de árboles superiores de Cacao (*Theobroma cacao* L.) seleccionados por el Programa de Mejoramiento Genético del CATIE. Tesis para optar al grado de Magíster Scientiae. Programa de enseñanza para el Desarrollo y la Conservación. Escuela de Postgrado. Centro Agronómico Tropical de Investigación y Enseñanza (CATIE). Turrialba, Costa Rica.
- AYCACHI, M. 2008. Caracterización Botánico Agronómica ex situ de 20 clones de cacao (*Theobroma cacao* L.) colectados en la cuenca del río Huallaga. Tesis para optar el título de Ingeniero Agrónomo. Facultad de Agronomía. Universidad Nacional Agraria de la Selva. Tingo – María. Huánuco.
- BARTRA, T. E. 1993. Caracterización Botánico Agronómica ex situ de 20 clones de cacao (*Theobroma cacao* L.) colectados en la cuenca del río Huallaga. Tesis para optar el título de Ingeniero Agrónomo. Facultad de Agronomía. Universidad Nacional Agraria de la Selva. Tingo – María. Huánuco.
- BORNET, B.; BRANCHARD, M. 2001. Nonanchored Inter Simple Sequence Repeat (ISSR) Markers: Reproducible and Specific Tools for Genome Fingerprinting. Commentary. Plant Molecular Biology Reporter. Volumen 19: pp: 209–215.
- BORRONE, J.W.; KUHN, D.N.; SCHNELL, R.J. 2004. Isolation, characterization, and development of WRKY genes as useful genetic markers in *Theobroma cacao*. Theoretical and Applied Genetics. Volumen 109 (3), pp. 495-507.
- BRAUDEAU, J. 1981. El Cacao. Blume Distribuidora, S.A. México, D.F. – México. 297 pp.
- BUNTJER, J.B. 1999. Cross Checker v. 2.9. Department of Plant Breeding. Wageningen University and Research Centre.  
[http://www.plantbreeding.wur.nl/UK/software\\_crosschecker.html](http://www.plantbreeding.wur.nl/UK/software_crosschecker.html)



- CHADHA, S.; GOPALAKRISHNA, T. 2007. Comparative assessment of REMAP and ISSR marker assays for genetic polymorphism studies in *Magnaporthe grisea*. Current Science, Volumen 93(5), Septiembre.
- CHARTERS, Y.M.; WILKINSON, M.J. 2000. The use of self-pollinated progenies as 'in-groups' for the genetic characterization of cocoa germplasm. Theor Appl Genet 100:160-166.
- CHEESMAN, E.E. 1944. Notes on the Nomenclature, classification and possible relationships of cocoa populations. Tropical Agriculture vol. XXI:144-159.
- CRISCI, J.; LÓPEZ, M. 1983. Introducción a la teoría y práctica de la taxonomía numérica. Serie n°26 OEA Washington D.C.
- DE VICENTE, M.C.; FULTON, T. 2003. Using molecular marker technology effectively in plant diversity studies. Vol 1. Learning module. CD-ROM. International Plant Genetic Resources Institute, Rome, Italy and Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA. <http://irc.igd.cornell.edu/MolecularMarkers/>
- DE VICENTE, M.C.; LÓPEZ, C.; FULTON, T. 2004. Genetic diversity analysis with molecular marker data. Vol 2. Learning module. CD-ROM. International Plant Genetic Resources Institute, Rome, Italy, Universidad Nacional Agraria 'La Molina', Peru, y Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA. <http://irc.igd.cornell.edu/MolecularMarkers/>
- DIAS, L.A.S. (ED.) 2001. Genetic Improvement of Cacao. Traducción del libro 'Melhoramento Genético do Cacaueiro' por C.E. Abreu-Reichart, Viçosa, M.G., Brazil; asistida por el Editor y la Food and Agriculture Organization (FAO). Versión electrónica Ecoport, por P. Griffiee (FAO). <http://ecoport.org/ep?SeachType=earticleView&earticleId=197>
- DIRECCIÓN GENERAL DE PROMOCIÓN AGRARIA (DGPA), 2007. Cacao-Chocolate. Ministerio de Agricultura. Año 1 - Boletín 1. Diciembre.
- DOS SANTOS, R.C.; PIRES, J. L.; VANDERLEY L., U.; PERES G. G., K.; BATISTA F., A.; DE CÁSSIA S. B., R. ; CRISTINE R., H. C. ; XAVIER C., R. ; AHNERT, D.; 2005. Assessment of genetic diversity on a sample of cocoa accessions resistant to witches' broom disease based on RAPD and pedigree data. Bragantia, Campinas. Volumen 64(3). pp. 361-368.
- DOYLE, J.J.; DOYLE, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19:11-15.

- EFOMBAGN, I.B.M.; MOTAMAYOR, J.C.; SOUNIGO, O.; ESKES, A.B.; NYASSÉ, S.; CILAS, C.; SCHNELL, R.; MANZANARES-DAULEUX, M.J.; KOLESNIKOVA-ALLEN, M. 2008. Genetic diversity and structure of farm and GenBank accessions of cacao (*Theobroma cacao* L.) in Cameroon revealed by microsatellite markers. *Tree Genetics & Genomes* 4:821–831. DOI 10.1007/s11295-008-0155-z
- ENGELS, J.M.M. 1986. The Systematic description of cacao clones and its significance for taxonomy and plant breeding. Doctoral thesis. Agricultural University, Wageningen. 125 pp.
- ENGELS, J.M.M.; VISSER, L. (Eds.). 2007. Guía para el manejo eficaz de un banco de germoplasma. Manuales para Bancos de Germoplasma No. 6. Bioversity International, Roma, Italia.
- EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50.
- FALEIRO, F.G. ; PIRES, J.L. ; MONTEIRO, W.R.; LOPES, U.V.; YAMADA, M.M.; PIEDRA, A.G.; MOURA, A.D. ; ARÉVALO-GARDINI, E.; MARQUES, J.R.B.; GRAMACHO, K.P.; GELAPE F., A.S. ; SANTOS, M.C.M. 2004. Variability in cacao accessions from the Brazilian, Ecuadorian, and Peruvian Amazons based on molecular markers. *Crop Breeding and Applied Biotechnology*. Volumen 4. pp.227-233. Brazilian Society of Plant Breeding.
- FERREIRA, M.E.; GRATTAPAGLIA, D. 1998. Introducción al Uso de Marcadores Moleculares en el Análisis Genético. 1<sup>ra</sup> Edición. EMBRAPA – CENARGEN Documento 20. Brasília D.F. – Brasil.
- GARCÍA, L.F. 2007. Identificación de los cultivares de cacao – Guía de campo. UNAS. USAID-PERÚ. ACDI/VOCA. Tingo María - Perú. 32pp.
- GARCÍA, L.F. 2000. Capítulo III. Grupos y variedades de cacao. En: Arca B., M. (Ed.) *El Cultivo del Cacao en la Amazonía Peruana*. Presidencia de la Republica. Ministerio de Agricultura, Lima - Perú.
- GARCÍA, E. 1998. Caracterización y Evaluación ex-situ de 21 clones de cacao (*Theobroma cacao* L.) de la colección Ucayali-Urubamba, en Tingo María. Tesis para optar el título de Ingeniero Agrónomo. Facultad de Agronomía. Universidad Nacional Agraria de la Selva. Tingo – María. Huánuco.

- GHISLAIN, M.; ZHANG, D.; FAJARDO, D.; HUAMAN, Z. ; HIJMANS, R.J. 1999. Marker assisted sampling of the cultivated Andean potato *Solanum phureja* collection using RAPD markers. Genetic Resources and Crop Evolution 46: 547-555.
- GONZÁLES, A.; AGUIRRE, X. 2007. Inter Simple Sequence Repeats (ISSR). Capítulo 19. pp. 567-571. En: Eguiarte, L. E.; Souza, V. ; Aguirre, X. (Compiladores). Ecología molecular. Secretaría de Medio Ambiente y Recursos Naturales. Instituto Nacional de Ecología. Universidad Nacional Autónoma de México. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. México.
- GONZALES, M.T. 1996. Caracterización Botánico Agronómica ex situ de 25 clones internacionales de cacao (*Theobroma cacao* L.). Tesis para optar el título de Ingeniero Agrónomo. Facultad de Agronomía. Universidad Nacional Agraria de la Selva. Tingo – María. Huánuco.
- GUERRERO, J.M. 2005. Estudio Taxonómico Intraespecífico de 48 clones de cacao (*Theobroma cacao* L.) de la colección Ucayali-Urubamba. Tesis para optar el título de Ingeniero Agrónomo. Facultad de Agronomía. Universidad Nacional Agraria de la Selva. Tingo – María. Huánuco.
- GUILTINAN, M.J. 2007. Cacao. En: Pua E.C.; Davey M.R. (Eds.) Biotechnology in Agriculture and Forestry. Transgenic Crops V. Section III.2 pp. 497-518. Springer-Verlag, Berlin Heidelberg.
- GUILTINAN, M.J.; VERICA, J.; ZHANG, D.; FIGUEIRA, A. 2008 Genomics of *Theobroma cacao*, “the Food of the Gods”. Chapter 6. pp. 145-170. En: Moore, P.H. ; Ming, R. (Eds.). Genomics of Tropical Crop Plants. Springer.
- INSTITUTO INTERAMERICANO DE COOPERACIÓN PARA LA AGRICULTURA (IICA). 2006. Protocolo estandarizado de oferta tecnológica para el cultivo del cacao en el Perú. USAID – WCF – Acceso. Noviembre 2006. Lima - Peru.
- INTERNATIONAL COCOA ORGANIZATION (ICCO), 2005. Inventory of the Health and Nutritional Attributes of Cocoa and Chocolate. Secretariat. Promotion Committee. Fifth meeting, December. Londres. 15 pp. <http://www.icco.org/documents/documents.aspx>
- INTERNATIONAL COCOA ORGANIZATION (ICCO), 2008. Annual Report 2006/07. London, United Kingdom. 48 pp. <http://www.icco.org/about/anualreport.aspx>

- JARAMILLO, S.; M. BAENA. 2000. Material de apoyo a la capacitación en conservación ex situ de recursos fitogenéticos. Instituto Internacional de Recursos Fitogenéticos (IPGRI), Cali, Colombia.
- KARP, A.; S. KRESOVICH; K.V. BHAT; W.G. AYAD; T. HODGIN 1997. Molecular tools in plant genetic resources conservation: a guide to the technologies. IPGRI technical bulletin n°2. International Plant Genetic Resources Institute, Rome, Italy. 47 pp.
- LACHENAUD, P.H. 1997. Genetic/Taxonomic Structuring of the *Theobroma cacao* L. species - Fresh Hypotheses. INGENIC Newsletter N°3 October. pp. 10-11.
- LACHENAUD, P.H.; PAULIN, D.; DUCAMP, M.; THEVENIN, J.-M. 2007. Twenty years of agronomic evaluation of wild cocoa trees (*Theobroma cacao* L.) from French Guyana. Scientia Horticulturae Volumen 113 pp. 313–321
- LÓPEZ, H. 1993. Caracterización Agronómico - Botánica "ex situ" de 20 clones de cacao (*Theobroma cacao* L.) recolectados de las cuencas de los ríos Ucayali y Urubamba. Tesis para optar el título de Ingeniero Agrónomo. Facultad de Agronomía. Universidad Nacional Agraria de la Selva. Tingo – María. Huánuco.
- MARITA, J. M.; NIENHUIS, J.; PIREX, J. L.; AITKEND, W. M. 2001. Analysis of Genetic Diversity in *Theobroma cacao* with Emphasis on Witches' Broom Disease Resistance. Crop Science. Volumen 41 pp. 1305-1316. Crop Science Society of America.
- MARTINEZ, W.J. 2007. Caracterización morfológica y molecular del Cacao Nacional Boliviano y de selecciones élites del Alto Beni, Bolivia. Tesis para optar al grado de Magíster Scientiae. Programa de enseñanza para el Desarrollo y la Conservación. Escuela de Postgrado. Centro Agronómico Tropical de Investigación y Enseñanza (CATIE). Turrialba, Costa Rica.
- M & O Consulting, 2008. Estudio de Caracterización del Potencial Genético del cacao en el Perú. Informe Final. Consultoría: 24/2007/PNRC/Lote 2. "Proyecto de Cooperación UE-Perú en materia de Asistencia Técnica relativa al Comercio - Apoyo al Programa Estratégico Nacional Exportaciones (PENX 2003-2013)". 136pp. <http://www.minag.gob.pe/direccion-general-de-competitividad-agraria/2.html>
- MOTAMAYOR, J.C.; LANAUD, C. 2002. Molecular Analysis of the Origin and Domestication of *Theobroma cacao* L. Chapter 8. pp 77-87. En: Engels,

- J.M.M.; Ramanatha Rao, V.; Brown A.H.D. ; Jackson M.T. (eds.). Managing Plant Genetic Diversity. IPGRI.
- MOTAMAYOR, J.C.; RISTERUCCI, A.M.; LOPEZ, P.A.; ORTIZ, C.F. MORENO, A.; LANAUD, C. 2002. Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity*. Volumen 89, pp. 380–386.
- MOTAMAYOR, J.C.; LACHENAUD, P.; DA SILVA E MOTA, J.W.; LOOR, R.; KUHN, D.N.; BROWN, J.S. ; SCHNELL, R.J. 2008. Geographic and genetic population differentiation of the amazonian chocolate tree (*Theobroma cacao* L.). *Plos ONE*. 2008; vol 3 (10): e3311. publicado en línea el 1 de octubre del 2008. Doi: 10.1371/journal.pone.0003311.  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2551746>
- N'GORAN, J.A.K.; LAURENT, V.; RISTERUCCI, A.M.; LANAUD, C. 1994. Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. *Heredity*. Volumen 73, pp. 589–597.
- OSORIO, M.E., SALAZAR, E. ZAMBRANO, A.Y. ; DEMEY, J.R. 2001. Estudio del cacao (*Theobroma cacao* L.) tipo criollo de la Costa central de Venezuela mediante amplificación aleatoria de ADN polimórfico. Póster 03-005. Resúmenes del IV Encuentro Latinoamericano de Biotecnología Vegetal. REDBIO. Goiania, Brasil. Del 4 al 8 de Junio del 2001.  
[http://www.redbio.org/portal/encuentros/enc\\_2001/posters/03/03pdf/03-005.pdf](http://www.redbio.org/portal/encuentros/enc_2001/posters/03/03pdf/03-005.pdf)
- OSORIO, M.E.; SALAZAR, E.; ZAMBRANO, A. 2003. Diversidad genética de una colección de cacao mediante RAPDs. *Agronomía Tropical* [online]. Vol. 53(1). Marzo. [http://www2.scielo.org.ve/scielo.php?script=sci\\_arttext&pid=S0002-192X2003000100001&lng=es&nrm=iso](http://www2.scielo.org.ve/scielo.php?script=sci_arttext&pid=S0002-192X2003000100001&lng=es&nrm=iso)
- PAGE, R.D.M. 2000. TREEVIEW (WIN32) 1.6.0.  
<http://taxonomy.zoology.gla.ac.uk/rod/rod.html>
- PANDURO C., W. E. 1998. Caracterización y Evaluación ex situ de 23 clones de cacao (*Theobroma cacao* L.) de la colección Ucayali, en Tingo María. Tesis para optar el título de Ingeniero Agrónomo. Facultad de Agronomía. Universidad Nacional Agraria de la Selva. Tingo – María. Huánuco.
- PARRA, P.; ORTIZ B., L.; GRAZIANI F., L. ; FIGUEROA R., R. 2001. Relaciones de similitud establecidas mediante el análisis de isoenzimas de los grupos de cacao criollo, trinitario y forastero de la zona de Cumboto, región centro

- norte costera del estado Aragua, Venezuela. *Agronomía Tropical* Volumen 51(4). pp. 575-592.
- PARRA, P.; GONZÁLEZ, M.; ORTIZ DE BERTORELLI, L.; GRAZIANI DE FARIÑAS, L.; FIGUEROA RUIZ, R. 2002. Caracterización de los tipos de cacaos criollo, trinitario y forastero de Cumboto, Aragua, mediante patrones electroforéticos de isoenzimas. *Agronomía Tropical* Volumen 52(3). pp. 277-297.
- PASQUALONE, A.; LOTTI, C.; BRUNO, A.; DE VITA, P.; DI FONZO, N. ; BLANCO A, 2000. Use of ISSR markers for cultivar identification in durum wheat. En: Royo, C.; Nachit, M.M.; Di Fonzo, N.; Araus, J.L. (Eds.) *Proceedings Durum wheat improvement in the Mediterranean region: new challenges*. Zaragoza (Spain), 12–14 April 2000. *Options Méditerranéennes, Serie: A, Séminaires Méditerranéens* volumen 40. pp. 157–161. <http://ressources.ciheam.org/om/pdf/a40/00600023.pdf>
- PIÑAN, A. 1993. Caracterización Botánico Agronómica ex situ de 12 clones internacionales de cacao (*Theobroma cacao* L.). Tesis para optar el título de Ingeniero Agrónomo. Facultad de Agronomía. Universidad Nacional Agraria de la Selva. Tingo – María. Huánuco.
- PROAMAZONIA (Programa para el Desarrollo de la Amazonía) y MINISTERIO DE AGRICULTURA 2003. Caracterización de las Zonas Productoras del Cacao en el Perú y su Competitividad. Informe Final. Lima – Perú.
- PROAMAZONIA (Programa para el Desarrollo de la Amazonía) y MINISTERIO DE AGRICULTURA 2004. Manual del Cultivo del Cacao. Lima – Perú.
- RENGIFO, L. P. 2002. Variabilidad Fenotípica y Heredabilidad de Caracteres Biométricos de la flor, fruto y semilla de 30 clones de cacao (*Theobroma cacao* L.) de distinto origen genético. Tesis para optar el título de Ingeniero Agrónomo. Facultad de Agronomía. Universidad Nacional Agraria de la Selva. Tingo – María. Huánuco.
- ROHLF, F.J. 2000. NTSYSpc Numerical Taxonomy and Multivariate Analysis System, Version 2.1. User Guide. Department of Ecology and Evolution. State University of New York, Stony Brook, NY.
- RONDÓN, J.B.; CUMANA C., L.J. 2005. Revisión Taxonómica del género *Theobroma* (Sterculiaceae) en Venezuela. *Acta Botánica Venezuelica*. Volumen 28 (1) Caracas.

- SARLA, N.; BOBBA, S.; SIDDIQ, E. A. 2003. ISSR and SSR markers based on AG and GA repeats delineate geographically diverse *Oryza nivara* accessions and reveal rare alleles. *Current Science*, Vol. 84, N° 5, 10
- SAAD, M. S.; RAO, V.R. (Eds.) 2001. Establishment and Management of Field Genebank, a Training Manual. IPGRI - Asia, the Pacific and Oceania, Serdang. Malaysia.
- SNEATH, P.H.A.; SOKAL, R.R. 1973. Numerical Taxonomy: the principles and practice of numerical classification. W.H. Freeman y Co. USA.
- SOKAL, R.; SNEATH, P. 1963. Principles of Numerical Taxonomy. Freeman, San Francisco.
- SORIA, J.; ENRIQUEZ, G.A. (Eds.) 1981. Catálogo Internacional de cultivares de cacao. Serie Técnica no. 6. Informe Técnico - Centro Agronómico Tropical de Investigación y Enseñanza (CATIE). Programa Cultivos Perennes. Turrialba (Costa Rica). 156 p.
- SOUNIGO, O.; UMAHARAN, R. CHRISTOPHER, Y. SANKAR, A.; RAMDAHIN, S. 2005. Assessing the genetic diversity in the International Cocoa Genebank, Trinidad (ICG,T) using isozyme electrophoresis and RAPD. *Genetic Resources and Crop Evolution*. Volumen 52. pp: 1111–1120. Springer.
- SPOONER, D.; VAN TREUREN, R.; DE VICENTE, M.C. 2005. Molecular markers for Genebank management. IPGRI Technical Bulletin No. 10. International Plant Genetic Resources Institute, Rome, Italy. 136 pp.
- SWANSON, J-D; LEE, A.C.; GUILTINAN, M. J. 2003. USDA Cacao DNA Fingerprinting Ring Test: Results from Penn State University. *INGENIC Newsletter* N° 8 July. pp. 22-24.
- TAKRAMA, J.F.; CERVANTES-MARTINEZ, C.; PHILLIPS-MORA, W.; BROWN, J. S.; MOTAMAYOR, J.C. ; SCHNELL, R.J. 2005. Determination of Off-Types in a Cocoa Breeding Programme using Microsatellites. *INGENIC Newsletter* N°10 September. pp 2-8.
- VIDAL, E. M. 1997. Caracterización ex situ de 16 clones de cacao (*Theobroma cacao* L.) de las cuenca del río Huallaga. Tesis para optar el título de Ingeniero Agrónomo. Facultad de Agronomía. Universidad Nacional Agraria de la Selva. Tingo – María. Huánuco.
- WARBURTON, M.; CROSSA, J. 2002. Data Analysis in the CIMMYT Applied Biotechnology Center. For Fingerprinting and Genetic Diversity Studies.

- WHITLOCK, B.A.; BAUM, D.A. 1999. Phylogenetic relationships of *Theobroma* and *Herrania* (Sterculiaceae) based on sequences of the nuclear gene *Vicilin*. *Systematic Botany*. Volumen 24 (2). pp 128-138. American Society of Plant Taxonomists.
- WOOD, G.A.R. 1975. *Cocoa*. 3ra edición. Longman Group Limited. Londres. 292 pp.
- YEH, F.C.; YANG, R.; BOYLE, T. 1999. POPGENE v1.31. University of Alberta and Centre for International Forestry Research. Canada.
- ZHANG, D.; ARÉVALO-GARDINI, E.; MISCHKE, S.; ZÚÑIGA-CERNADES, L.; BARRETO-CHÁVEZ, A.; ADRIAZOLA A., J. 2006. Genetic Diversity and Structure of Managed and Semi-natural Populations of *Cocoa* (*Theobroma cacao*) in the Huallaga and Ucayali Valleys of Peru. *Annals of Botany* 10.1093/aob/mcl146.



**ANEXO N° 1**

Datos morfológicos más resaltantes recopilados de los clones e híbridos de cacao analizados. N.D.: no determinado.

<b>Clones Nacionales</b>					
<b>código</b>	<b>color de fruto inmaduro</b>	<b>forma de fruto</b>	<b>rugosidad del fruto</b>	<b>número ovulos por ovario</b>	<b>color de cotiledón</b>
C85	N.D.	N.D.	N.D.	N.D.	N.D.
ChuC	verde	elíptico	intermedio	N.D.	violeta
Seño	verde	oblongo	ligero	N.D.	-
SeñoA	verde	oblongo	fuerte	N.D.	-
U54	verde o verde intermedio	ovalado	rugoso o intermedio	40	violeta
U60	verde claro	ovalado	rugoso	42	morado
U68	verde o verde intermedio	ovalado	liso	43	morado
U70	verde o verde intermedio	ovalado	liso	33	blanco
H12	verde claro o verde	ovalado o elíptico	intermedio o ligero	60	violeta
H37	verde	elíptico	intermedio	N.D.	morado
H60	verde	elíptico	ligero	N.D.	morado
H61	verde	elíptico	ligero	N.D.	morado
SilvHu	verde	elíptico	ligero	N.D.	violeta
ChuS	verde	elíptico	fuerte	N.D.	marron
IMC67	verde	ovoide o elíptico	ligero	63	morado
P7	verde	ovoide o elíptico	ligero	54	morado
SCA6	verde	elíptico u ovoide	ligero	45	morado
PA150	verde	elíptico	intermedio	48	morado
C42	N.D.	N.D.	N.D.	N.D.	N.D.
C53	rojo	oblongo	ligero	N.D.	morado
C65	rojo	redondeado	ligero	N.D.	morado
C69	rojo	redondeado	ligero	N.D.	morado
C81	verde	oblonga	intermedio	N.D.	morado

Clones Nacionales continúa...

Clones Nacionales					
código	color de fruto inmaduro	forma de fruto	rugosidad del fruto	número ovulos por ovario	color de cotiledón
U1	verde	elíptico	intermedio	45	violeta
U9	verde	ovalado o elíptico	intermedio o ligero	37	violeta
U12	verde pigmentado	ovoide	ligero	51	morado
U15	verde con pigmentación	ovalado o elíptico	rugoso o intenso rugoso	38	violeta
U26	verde	ovalado u ovoide	rugoso o intermedio	48	morado
U43	verde intermedio	ovalado	ligero	40	morado
U48	verde o verde intermedio	ovalado	intermedio o liso	37	violeta

Clones Internacionales					
código	color de fruto inmaduro	forma de fruto	rugosidad del fruto	número ovulos por ovario	color de cotiledón
CAT4	verde	esférico	liso	48	blanco
Pand	verde	elíptico	ligero	54	violeta
CCN51	rojo	elíptico	fuertemente rugoso	57	morado
EET228	verde	ovoide	intermedio	51	violeta
EET233	verde pigmentado	abovado	ligero	56	morado
EET400	verde	elíptico	rugoso	46	morado
EET62	verde pigmentado	elíptico	rugoso	58	violeta
ICS1	rojo	elíptico	ligero	42	morado
ICS6	verde	elíptico	ligero	48	morado
ICS39	verde pigmentado	oblongo	intermedio	44	violeta
ICS78	rojo	elíptico	intermedio	43	violeta
ICS95	rojo	oblongo	intermedio	42	morado
UF29	verde	ovoide	ligero	47	morado
UF667	rojo grisáceo	abovado	intermedio	42	morado

Híbridos Proyecto Bioversity/UNAS					
código	color de fruto inmaduro	forma de fruto	rugosidad del fruto	número ovulos por ovario	color de cotiledón
HTM1	rojo	elíptico	ligero	-	violeta
I12.12	rojo	elíptico	ligero	-	morado
I14.20	rojo	elíptico	intermedio	-	morado
I16.20	rojo	elíptico	intermedio	-	morado
M1.7	verde	elíptico	ligero	-	morado
M18.16	verde pigmentado	elíptico	intermedio	-	morado

**ANEXO N° 2**

Lista de algunos caracteres agronómicos de importancia recopilados de los clones e híbridos de cacao analizados. NUSE: número de semillas, PESE: peso de semilla. N.D.: no determinado.

<b>Clones Nacionales</b>					
<b>código</b>	<b>Compatibilidad</b>	<b>% de grasa en semilla</b>	<b>Índice de mazorca</b>	<b>NUSE</b>	<b>PESE (g)</b>
C85	N.D.	N.D.	N.D.	N.D.	N.D.
ChuC	N.D.	N.D.	N.D.	N.D.	N.D.
Seño	N.D.	N.D.	N.D.	N.D.	N.D.
SeñoA	N.D.	N.D.	N.D.	N.D.	N.D.
U54	autoincompatible	49	33	38	1.93
U60	autoincompatible	58	22	35	2.18
U68	autocompatible	50	27	28.35	0.91
U70	autoincompatible	52	37	31	0.88
H12	autoincompatible	40,31	19	47.35	1.15
H37	autocompatible	53	23	47	0.9
H60	N.D.	N.D.	N.D.	N.D.	0.8
H61	N.D.	N.D.	N.D.	N.D.	N.D.
SilvHu	autoincompatible	N.D.	N.D.	N.D.	N.D.
ChuS	N.D.	N.D.	N.D.	N.D.	N.D.
IMC67	autoincompatible	48	16.36	47	1.1
P7	autoincompatible	48	21.94	34	1.0
SCA6	autoincompatible	59	29.9	38	0.88
PA150	autoincompatible	N.D.	18	44	1.26
C42	N.D.	N.D.	N.D.	N.D.	N.D.
C53	N.D.	N.D.	N.D.	N.D.	N.D.
C65	N.D.	N.D.	N.D.	N.D.	N.D.
C69	N.D.	N.D.	N.D.	N.D.	N.D.
C81	N.D.	N.D.	N.D.	N.D.	N.D.
U1	autoincompatible	53	36	39	0.74
U9	autocompatible	58	28	34	1.10
U12	autocompatible	55	15	45	1.15
U15	autoincompatible	60	32	38	0.81
U26	autoincompatible	59	17	39	1.40
U43	autoincompatible	49	24	28	1.22
U48	autoincompatible	50	17	38	1.51

<b>Clones Internacionales</b>					
<b>código</b>	<b>Compatibilidad</b>	<b>% de grasa en semilla</b>	<b>Índice de mazorca</b>	<b>NUSE</b>	<b>PESE (g)</b>
CAT4	autocompatible	N.D.	20.3	41	1.22
Pand	autocompatible	N.D.	16.9	38	1.41
CCN51	autocompatible	54	16	45	1.4
EET228	autocompatible	51	22	38	1.77
EET233	autoincompatible	N.D.	20.3	41	1.22
EET400	autoincompatible	49	19	32	1.51
EET62	autoincompatible	N.D.	14.5	38	1.81
ICS1	autocompatible	51	23	36	1.84
ICS6	autocompatible	54	16	39	2.30
ICS39	autoincompatible	51	13	34	2.62
ICS78	autocompatible	49	12.15	34	2.42
ICS95	autocompatible	56	23.93	30	1.39
UF29	autocompatible	57	20.39	38	1.29
UF667	autoincompatible	45	15.30	30	2.19

<b>Híbridos Proyecto Bioversity/UNAS</b>					
<b>código</b>	<b>Compatibilidad</b>	<b>% de grasa en semilla</b>	<b>Índice de mazorca</b>	<b>NUSE</b>	<b>PESE (g)</b>
HTM1	autocompatible	N.D.	N.D.	N.D.	N.D.
I12.12	autocompatible	N.D.	24	35	1.2
I14.20	autocompatible	N.D.	19	37	1.4
I16.20	autocompatible	N.D.	21	39	1.4
M1.7	autocompatible	N.D.	N.D.	N.D.	1.6
M18.16	autoincompatible	N.D.	N.D.	N.D.	1.4

**ANEXO N° 3**

Datos de Pureza ( $A_{260}/A_{280}$ ) y concentraciones de ADN (ng/ $\mu$ l) para las 50 accesiones de cacao. Lecturas obtenidas del biofotómetro Eppendorf.

genotipos	Relación $A_{260}/A_{280}$	Concentración ADN (ng/ $\mu$ l)
EET400	1,33	193
EET228	1,64	591
EET233	1,29	449
EET62	1,45	228
ICS6	1,21	380
ICS78	1,36	535
ICS1	1,56	630
ICS95	1,66	111
ICS39	1,48	678
M18.16	1,39	211
M1.7	1,22	1802
UF29	1,65	198
UF667	1,48	586
U1	2,04	297
CAT4	1,43	301
U9	1,54	377
U12	1,60	389
U15	1,31	928
U26	1,47	365
U43	1,20	383
U48	1,49	452
U54	1,33	450
U60	1,38	1388
U68	1,31	188
U70	1,31	139
SCA6	2,06	248
C42	1,69	413
C53	1,88	600
C65	1,53	573
HTM1	1,25	406
CCN51	1,61	350
IMC67	1,70	299
I12.12	1,64	508
I14.20	1,35	353
I16.20	1,51	529
C69	1,64	387

Continua...

<b>genotipos</b>	<b>Relación <math>A_{260}/A_{280}</math></b>	<b>Concentración ADN (ng/μl)</b>
C81	1,52	362
C85	1,99	282
SilvHu	1,53	487
H12	1,20	1078
H37	1,40	317
H61	1,81	180
H60	1,60	401
Pand	1,57	439
PA150	1,39	444
P7	1,32	303
ChuC	1,40	219
ChuS	2,39	111
SeñoA	1,69	476
Seño	1,97	376

#### ANEXO N° 4

Lista de los 59 Iniciadores ISSR. Los códigos corresponden a iniciadores de la Universidad de British Columbia. Los códigos que resaltados en negrita pertenecen a los 13 iniciadores cuyos perfiles de amplificación fueron más consistentes que del resto.

Iniciadores anclados			Iniciadores no anclados		
N°	código	Secuencia	N°	código	Secuencia
1	BOR 1	GAGC(CAA) <sub>5</sub>	21	UBC826	(AC) <sub>8</sub> C
2	BOR 2	CTG(AG) <sub>8</sub>	22	UBC827	(AC) <sub>8</sub> G
3	BOR 3	(AG) <sub>8</sub> TG	<b>23</b>	<b>UBC829</b>	<b>(TG)<sub>8</sub>C</b>
4	UBC807	(AG) <sub>8</sub> T	<b>24</b>	<b>UBC830</b>	<b>(TG)<sub>8</sub>G</b>
5	UBC808	(AG) <sub>8</sub> C	<b>25</b>	<b>UBC834</b>	<b>(AG)<sub>8</sub>YT</b>
6	UBC809	(AG) <sub>8</sub> G	<b>26</b>	<b>UBC835</b>	<b>(AG)<sub>8</sub>YC</b>
7	UBC810	(GA) <sub>8</sub> T	<b>27</b>	<b>UBC841</b>	<b>(GA)<sub>8</sub>YC</b>
8	UBC811	(GA) <sub>8</sub> C	<b>28</b>	<b>UBC842</b>	<b>(GA)<sub>8</sub>YG</b>
9	UBC812	(GA) <sub>8</sub> A	29	UBC844	(CT) <sub>8</sub> RC
10	UBC813	(CT) <sub>8</sub> T	30	UBC848	(CA) <sub>8</sub> RG
11	UBC814	(CT) <sub>8</sub> A	31	UBC849	(GT) <sub>8</sub> TYC
12	UBC815	(CT) <sub>8</sub> G	32	UBC858	(TG) <sub>8</sub> RT
13	UBC816	(CA) <sub>8</sub> T	33	UBC859	(TG) <sub>8</sub> RC
14	UBC817	(CA) <sub>8</sub> A	34	UBC860	(TG) <sub>8</sub> RA
15	UBC818	(CA) <sub>8</sub> G	<b>35</b>	<b>UBC884</b>	<b>HBH(AG)<sub>7</sub></b>
16	UBC820	(GT) <sub>8</sub> C	36	UBC885	BHB(GA) <sub>7</sub>
17	UBC822	(TC) <sub>8</sub> A	37	UBC887	DVD(TC) <sub>7</sub>
18	UBC823	(TC) <sub>8</sub> G	38	UBC889	DBD(AC) <sub>7</sub>
19	UBC824	(TC) <sub>8</sub> G	39	UBC890	VHV(GT) <sub>8</sub>
20	UBC825	(AC) <sub>8</sub> T	40	UBC891	HVH(TG) <sub>9</sub>
			<b>41</b>	<b>BOR 5</b>	<b>(AG)<sub>8</sub></b>
			42	BOR 6	(CCA) <sub>5</sub>
			43	BOR7	(AC) <sub>8</sub>
			<b>44</b>	<b>GUP-1</b>	<b>(ACTG)<sub>4</sub></b>
			<b>45</b>	<b>GUP 2</b>	<b>(GACA)<sub>4</sub></b>
			46	GUP3	(GATA) <sub>4</sub>
			47	GUP 4	(GACAC) <sub>4</sub>
			48	UBC861	(ACC) <sub>6</sub>
			49	UBC862	(ACG) <sub>6</sub>
			<b>50</b>	<b>UBC864</b>	<b>(ATG)<sub>6</sub></b>
			51	UBC865	(CCG) <sub>6</sub>
			52	UBC866	(CTC) <sub>6</sub>
			53	UBC867	(GGC) <sub>6</sub>
			<b>54</b>	<b>UBC868</b>	<b>(GAA)<sub>6</sub></b>
			55	UBC876	(GATA) <sub>2</sub> (GACA) <sub>2</sub>
			56	UBC878	(GGAT) <sub>4</sub>
			57	UBC879	(CTTCA) <sub>3</sub>
			<b>58</b>	<b>UBC880</b>	<b>(GGAGA)<sub>3</sub></b>
			59	UBC881	(GGGTG) <sub>3</sub>

**Nota:** los iniciadores degenerados son mezclas de iniciadores PCR donde los símbolos en la respectiva fórmula representan: **B** = G, T, ó C; **D** = G, A, ó T; **H** = A, T, ó C; **V** = G, A, ó C; **R** = A, G e **Y** = C, T.



## ANEXO N° 5

Datos Binarios (presencia: 1 ó ausencia: 0) de 73 loci (caracteres) para 50 genotipos (OTUs). El código 2 es el asignado a los datos perdidos o “missing data”. Se muestran sólo los loci con un porcentaje de datos perdidos menor o igual al 10%.

	loci	EET400	EET228	EET233	EET62	ICS6	ICS78	ICS1	ICS95	ICS39	M18.16	M1.7	UF29	UF667	U1	CAT4	U9	U12	U15	U26	U43	U48	U54	U60	U68	U70	SCA6	C42	C53	C65	HTM1	CCN51	IMC67	I12.12	I14.20	I16.20	C69	C81	C85	SilvHu	H12	H37	H61	H60	Pand	PA150	P7	ChuC	ChuS	SeñoA	Seño	
UBC834	L1	2	1	1	1	1	2	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	2				
UBC834	L2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1				
UBC834	L5	2	1	1	1	1	2	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0				
UBC834	L7	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2	0	0			
UBC834	L11	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2				
UBC834	L12	1	1	1	1	1	0	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1				
UBC834	L14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0				
UBC834	L18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1				
UBC834	L19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
UBC834	L20	1	2	1	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	2		
UBC834	L21	1	2	2	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0		
UBC835	L1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0		
UBC835	L2	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
UBC835	L3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	1	
UBC835	L4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
UBC835	L5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
UBC835	L6	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
UBC835	L9	0	2	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	2	2	1	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
UBC835	L10	0	0	0	0	0	0	0	2	2	2	0	0	1	0	0	0	0	2	2	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
UBC835	L20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0		
UBC835	L22	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1		
UBC835	L23	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
UBC835	L26	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

loci		EET400	EET228	EET233	EET62	ICS6	ICS78	ICS1	ICS95	ICS39	M18.16	M1.7	UF29	UF667	U1	CAT4	U9	U12	U15	U26	U43	U48	U54	U60	U68	U70	SCA6	C42	C53	C65	HTM1	CCN51	IMC67	I12.12	I14.20	I16.20	C69	C81	C85	SilvHu	H12	H37	H61	H60	Pand	PA150	P7	ChuC	ChuS	SeñoA	Seño			
UBC835	L27	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
UBC835	L28	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1			
UBC841	L5	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0		
UBC841	L8	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	2	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0		
UBC841	L9	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	2	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
UBC841	L10	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	2	1	1	1	1	0			
UBC841	L13	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	1		
UBC841	L16	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	2	1	1	0	0	1	0	0	0	0	1	0	0	0	0	2	2	0	2	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	2	0
UBC841	L17	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	1	1	0	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	1	0	0	1	1	0	1	1	0	1	0	1	0	1	2	0	0	0	
UBC841	L24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		
UBC841	L25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
UBC841	L26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
UBC841	L28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	
UBC841	L29	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	
UBC841	L33	0	0	2	0	0	2	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	0	1	0	0	0	0	0	1	1	
UBC841	L34	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	2	0	1		
UBC841	L36	0	0	0	0	0	2	0	2	0	0	2	0	0	0	0	0	0	0	2	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UBC841	L37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
UBC841	L40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
UBC841	L42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	
UBC841	L45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
UBC842	L4	1	1	1	0	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1		
UBC842	L10	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	1	1	1	1	2	2		
UBC842	L11	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	0	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	2			
UBC842	L12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1		
UBC842	L13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	0	0	0	1	1	1	0	0	0	0	1	0	1		
UBC842	L14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1		

loci		EET400	EET228	EET233	EET62	ICS6	ICS78	ICS1	ICS95	ICS39	M18.16	M1.7	UF29	UF667	U1	CAT4	U9	U12	U15	U26	U43	U48	U54	U60	U68	U70	SCA6	C42	C53	C65	HTM1	CCN51	IMC67	I12.12	I14.20	I16.20	C69	C81	C85	SilvHu	H12	H37	H61	H60	Pand	PA150	P7	ChuC	ChuS	SeñoA	Seño		
UBC842	L15	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	2	1	2	1	0	1	0	1	0	1		
UBC842	L16	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	0	1	1	1	0	1	1	0			
UBC842	L20	2	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1			
UBC842	L22	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	2	0	0	1	0	1	0	0	0	0		
UBC842	L24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	0	2		
UBC842	L25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1			
UBC842	L27	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2				
UBC842	L28	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	1	0	0	1	0	0	0	0	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	2				
UBC842	L29	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	2	0	1	1	1	0	1	0	1	0	1	0	1	0	0	0	0	2	1	1	1	0	0	0	0	2	0	1			
UBC842	L30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
UBC884	L4	1	1	1	1	1	1	1	1	1	1	1	0	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1			
UBC884	L5	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0				
UBC884	L10	0	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	2	1	1	0	0	0	1	1	2	0	2	1	1	0	0	1	1	1		
UBC884	L11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1			
UBC884	L12	0	0	0	0	1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	2	0	0	2	0	1	0	1	0	1	0	0	0	1		
UBC884	L13	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0		
UBC884	L14	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	2		
UBC884	L16	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	0		
UBC884	L17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
UBC884	L18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
UBC884	L20	0	0	0	0	0	2	0	0	0	0	0	0	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	1	1	1	1	1	0	1	1	1	1	1	2	0		
UBC884	L21	0	1	1	1	1	2	1	0	0	1	0	1	1	1	0	2	2	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1		
UBC884	L22	1	1	1	1	1	1	0	1	0	1	0	2	0	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	2	0	2	1	0	0	0	0	1	2

## ANEXO N° 6

Valores descriptivos calculados para cada locus. Loci con un porcentaje de datos perdidos  $\leq 10\%$ . Mono = locus monomórfico.

	N° de presencias	N° de ausencias	N° de missing data	% missing data por locus	N° de loci verdaderos	P (freq genotípica)	Q (freq genotípica)	p (freq alelica presencias)	q (freq alelica ausencias)	PIC=1-p <sup>2</sup> -q <sup>2</sup>	Índice del marcador ISSR	loci monomórfico (mono) o raro	genotipo con loci raro
UBC834 L1	44	1	5	10	45	0,98	0,02	0,851	0,149	0,25			
UBC834 L2	48	0	2	4	48	1,00	0,00	1,000	0,000	0,00		mono	
UBC834 L5	43	5	2	4	48	0,90	0,10	0,677	0,323	0,44			
UBC834 L7	1	46	3	6	47	0,02	0,98	0,011	0,989	0,02		raro	U1
UBC834 L11	47	0	3	6	47	1,00	0,00	1,000	0,000	0,00	2,07	mono	
UBC834 L12	48	1	1	2	49	0,98	0,02	0,857	0,143	0,24			
UBC834 L14	47	2	1	2	49	0,96	0,04	0,798	0,202	0,32			
UBC834 L18	49	0	1	2	49	1,00	0,00	1,000	0,000	0,00		mono	
UBC834 L19	50	0	0	0	50	1,00	0,00	1,000	0,000	0,00		mono	
UBC834 L20	44	4	2	4	48	0,92	0,08	0,711	0,289	0,41			
UBC834 L21	45	3	2	4	48	0,94	0,06	0,750	0,250	0,38			
UBC835 L1	6	44	0	0	50	0,12	0,88	0,062	0,938	0,12			
UBC835 L2	6	43	1	2	49	0,12	0,88	0,063	0,937	0,12			
UBC835 L3	5	44	1	2	49	0,10	0,90	0,052	0,948	0,10			
UBC835 L4	1	45	4	8	46	0,02	0,98	0,011	0,989	0,02		raro	Seño
UBC835 L5	1	46	3	6	47	0,02	0,98	0,011	0,989	0,02		raro	Seño
UBC835 L6	1	48	1	2	49	0,02	0,98	0,010	0,990	0,02		raro	Seño
UBC835 L9	12	34	4	8	46	0,26	0,74	0,140	0,860	0,24	1,63		
UBC835 L10	6	39	5	10	45	0,13	0,87	0,069	0,931	0,13			
UBC835 L20	1	49	0	0	50	0,02	0,98	0,010	0,990	0,02		raro	C81
UBC835 L22	45	0	5	10	45	1,00	0,00	1,000	0,000	0,00		mono	
UBC835 L23	3	45	2	4	48	0,06	0,94	0,032	0,968	0,06		raro	Pand
UBC835 L26	7	43	0	0	50	0,14	0,86	0,073	0,927	0,13			
UBC835 L27	49	1	0	0	50	0,98	0,02	0,859	0,141	0,24			
UBC835 L28	45	4	1	2	49	0,92	0,08	0,714	0,286	0,41			
UBC841 L5	1	44	5	10	45	0,02	0,98	0,011	0,989	0,02		raro	CAT4
UBC841 L8	11	37	2	4	48	0,23	0,77	0,122	0,878	0,21			
UBC841 L9	44	5	1	2	49	0,90	0,10	0,681	0,319	0,43			
UBC841 L10	44	4	2	4	48	0,92	0,08	0,711	0,289	0,41			
UBC841 L13	10	39	1	2	49	0,20	0,80	0,108	0,892	0,19			
UBC841 L16	9	36	5	10	45	0,20	0,80	0,106	0,894	0,19			
UBC841 L17	18	29	3	6	47	0,38	0,62	0,214	0,786	0,34			
UBC841 L24	1	49	0	0	50	0,02	0,98	0,010	0,990	0,02		raro	Seño
UBC841 L25	1	48	1	2	49	0,02	0,98	0,010	0,990	0,02	2,43	raro	H37
UBC841 L26	1	49	0	0	50	0,02	0,98	0,010	0,990	0,02		raro	H60
UBC841 L28	1	47	2	4	48	0,02	0,98	0,010	0,990	0,02		raro	Seño
UBC841 L29	4	45	1	2	49	0,08	0,92	0,042	0,958	0,08		raro	I16.20,H37,H60
UBC841 L33	9	37	4	8	46	0,20	0,80	0,103	0,897	0,19			
UBC841 L34	2	44	4	8	46	0,04	0,96	0,022	0,978	0,04		raro	H61, Seño
UBC841 L36	1	44	5	10	45	0,02	0,98	0,011	0,989	0,02		raro	U68
UBC841 L37	6	42	2	4	48	0,13	0,88	0,065	0,935	0,12			
UBC841 L40	1	47	2	4	48	0,02	0,98	0,010	0,990	0,02		raro	ChuS
UBC841 L42	2	47	1	2	49	0,04	0,96	0,021	0,979	0,04		raro	ChuS, Seño
UBC841 L45	2	48	0	0	50	0,04	0,96	0,020	0,980	0,04		raro	U26, U68

## ANEXO N° 6 - Parte 2

	N° de presencias	N° de ausencias	N° de missing data	% missing data por locus	N° de loci verdaderos	P (freq genotípica)	Q (freq genotípica)	p (freq alelica presencias)	q (freq alelica ausencias)	PIC=1-p <sup>2</sup> -q <sup>2</sup>	Índice del marcador ISSR	loci monomórfico (mono) o raro	genotipo con loci raro
UBC842 L4	44	6	0	0	50	0,88	0,12	0,654	0,346	0,45			
UBC842 L10	41	7	2	4	48	0,85	0,15	0,618	0,382	0,47			
UBC842 L11	44	1	5	10	45	0,98	0,02	0,851	0,149	0,25			
UBC842 L12	46	2	2	4	48	0,96	0,04	0,796	0,204	0,32			
UBC842 L13	37	11	2	4	48	0,77	0,23	0,521	0,479	0,50			
UBC842 L14	46	3	1	2	49	0,94	0,06	0,753	0,247	0,37			
UBC842 L15	42	5	3	6	47	0,89	0,11	0,674	0,326	0,44			
UBC842 L16	44	4	2	4	48	0,92	0,08	0,711	0,289	0,41	4,56		
UBC842 L20	11	38	1	2	49	0,22	0,78	0,119	0,881	0,21			
UBC842 L22	14	32	4	8	46	0,30	0,70	0,166	0,834	0,28			
UBC842 L24	1	47	2	4	48	0,02	0,98	0,010	0,990	0,02		raro	ChuS
UBC842 L25	4	46	0	0	50	0,08	0,92	0,041	0,959	0,08		raro	U26,U48,U54,Seño
UBC842 L27	3	46	1	2	49	0,06	0,94	0,031	0,969	0,06		raro	EET228,U54,ChunC
UBC842 L28	11	36	3	6	47	0,23	0,77	0,125	0,875	0,22			
UBC842 L29	28	18	4	8	46	0,61	0,39	0,374	0,626	0,47			
UBC842 L30	50	0	0	0	50	1,00	0,00	1,000	0,000	0,00		mono	
UBC884 L4	46	2	2	4	48	0,96	0,04	0,796	0,204	0,32			
UBC884 L5	45	5	0	0	50	0,90	0,10	0,684	0,316	0,43			
UBC884 L10	38	8	4	8	46	0,83	0,17	0,583	0,417	0,49			
UBC884 L11	48	2	0	0	50	0,96	0,04	0,800	0,200	0,32			
UBC884 L12	16	32	2	4	48	0,33	0,67	0,184	0,816	0,30			
UBC884 L13	11	39	0	0	50	0,22	0,78	0,117	0,883	0,21			
UBC884 L14	46	3	1	2	49	0,94	0,06	0,753	0,247	0,37	4,20		
UBC884 L16	46	2	2	4	48	0,96	0,04	0,796	0,204	0,32			
UBC884 L17	49	0	1	2	49	1,00	0,00	1,000	0,000	0,00		mono	
UBC884 L18	2	47	1	2	49	0,04	0,96	0,021	0,979	0,04		raro	U43, U70
UBC884 L20	26	22	2	4	48	0,54	0,46	0,323	0,677	0,44			
UBC884 L21	34	13	3	6	47	0,72	0,28	0,474	0,526	0,50			
UBC884 L22	27	19	4	8	46	0,59	0,41	0,357	0,643	0,46			

# **ANEXO N° 7**

Correspondencia entre los 5 iniciadores utilizados y los genotipos que muestran fragmentos amplificados con una frecuencia alélica ( $p$ ) menor a 0,05 (indicados en el Anexo N° 6).

Iniciador	Ucayali-Urubamba							Huallaga			Colectas Recientes				Internacionales			Híbridos
	U1	U26	U48	U43	U54	U68	U70	H37	H60	H61	Seño	C81	ChuS	ChuC	EET228	Pand	CAT4	I16.20
UBC834	X																	
UBC835											XXX	X				X		
UBC841		X				XX		XX	XX	X	XXXX		XX				X	X
UBC842		X	X		XX						X		X	X	X			
UBC884				X			X											

## ANEXO Nº 8

Archivos Lote o "*Batch files*" del programa NTSYSpc v2.1p usados para el análisis multivariado de los datos moleculares.

1. Agrupamiento UPGMA con la opción SAHN, Coeficiente de similitud *Jaccard* (J). Además se realiza la comparación de matrices de distancia y cofenéticas.

```
"calcular las distancias entre las OTUs como datos cualitativos
(binarios)
*simqual o=datos.nts r=dist.nts c=j
"realizar el agrupamiento de datos con analisis UPGMA
*sahn o=dist.nts r=tree.nts cm=upgma
"mostrar fenograma
*tree o=tree.nts
"calcular la matriz cofenética y el coeficiente de correlación
*coph o=tree.nts r=coph.nts
"graficar la correlación de matrices
*mxcomp x=coph.nts y=dist.nts
```

2. Análisis de Coordenadas Principales.

```
"calcular las distancias entre objetos
*simqual o= datos.nts r=dist.nts c=j
" realizar el Double-center de la matriz de distancia
*dcenter o=dist.nts r=dcent.nts
"calcular los eigenvectores que corresponden a las proyecciones de
objetos (dos dimensiones)
*eigen o=dcent.nts n=2 r=proj.nts val=eval.nts
*output o=proj.nts
" Mostrar gráfico - la direccion de los datos es "col" (columnas)
*mxplot o=proj.nts d=col
```

3. Escalamiento Multidimensional no métrico.

```
"calcular las distancias entre objetos
*simqual o= datos.nts r=dist.nts c=j
" realizar el Double-center de la matriz de distancia
*dcenter o=dist.nts r=dcent.nts
"calcular los eigenvectores que corresponden a las proyecciones de
objetos (dos dimensiones)
*eigen o=dcent.nts n=2 r=init.nts
"Escalamiento multidimensional no métrico usando la solución
inicial
*MDSCALE O=dist.nts n=2 d=col I=init.nts R=final.nts
"Graficar
*MXPLOT O=final.nts
```

**ANEXO N° 9**

Matriz de similitud de la comparación uno a uno entre las 50 OTUs. Coeficiente de Asociación *Jaccard*.

## Matriz de Similitud - Parte 1

	EET400	EET228	EET233	EET62	ICS6	ICS78	ICS1	ICS95	ICS39	M18.16	M1.7	UF29	UF667	U1	CAT4	U9	U12
EET400	1.000																
EET228	0.8065	1.000															
EET233	0.8333	0.9697	1.000														
EET62	0.8387	0.9091	0.9375	1.000													
ICS6	0.8125	0.9143	0.9118	0.9118	1.000												
ICS78	0.8333	0.8710	0.9000	0.9000	0.9355	1.000											
ICS1	0.7353	0.8889	0.8857	0.8333	0.9167	0.8485	1.000										
ICS95	0.8438	0.8889	0.9143	0.8611	0.8919	0.9091	0.8684	1.000									
ICS39	0.8387	0.8571	0.8824	0.8286	0.8611	0.8750	0.8889	0.9167	1.000								
M18.16	0.8438	0.9143	0.9412	0.8857	0.8649	0.8485	0.8421	0.8684	0.8889	1.000							
M1.7	0.8667	0.8824	0.9091	0.8529	0.8333	0.8438	0.8611	0.8889	0.9697	0.9143	1.000						
UF29	0.4000	0.4706	0.4545	0.4242	0.4706	0.4516	0.5294	0.4857	0.4706	0.4571	0.4412	1.000					
UF667	0.7647	0.8378	0.8611	0.8108	0.8421	0.8235	0.8684	0.8684	0.9429	0.8919	0.8889	0.4722	1.000				
U1	0.6341	0.7209	0.7209	0.6818	0.7111	0.7000	0.7333	0.7174	0.7273	0.7727	0.7045	0.4048	0.7727	1.000			
CAT4	0.7059	0.7568	0.7568	0.7568	0.7000	0.6944	0.7250	0.7500	0.8108	0.8158	0.8333	0.4054	0.8158	0.7674	1.000		
U9	0.6944	0.7632	0.7632	0.7632	0.7073	0.7027	0.7317	0.7381	0.7692	0.7750	0.7692	0.4324	0.7750	0.8372	0.8462	1.000	
U12	0.7500	0.8158	0.8158	0.7692	0.7561	0.7568	0.7805	0.7857	0.8205	0.8250	0.8205	0.4211	0.8250	0.8409	0.8462	0.9487	1.000
U15	0.6944	0.7436	0.7895	0.7000	0.6905	0.7027	0.7143	0.7381	0.7500	0.8000	0.7692	0.3590	0.7317	0.8372	0.7949	0.8462	0.8537
U26	0.6667	0.7561	0.8000	0.7143	0.7045	0.6750	0.6889	0.7111	0.7381	0.8095	0.7381	0.3571	0.7442	0.7500	0.7619	0.8095	0.7955
U43	0.6000	0.7027	0.7297	0.6579	0.6500	0.6286	0.6341	0.6750	0.7027	0.7838	0.6842	0.3243	0.7179	0.7442	0.6842	0.7632	0.7692
U48	0.5625	0.6765	0.7059	0.6286	0.6216	0.5938	0.6053	0.6486	0.6765	0.7143	0.6765	0.3143	0.6944	0.7000	0.6111	0.6944	0.7500
U54	0.6500	0.7619	0.7442	0.6818	0.6739	0.6585	0.6596	0.7174	0.7045	0.7727	0.7209	0.3721	0.7174	0.7347	0.7273	0.7727	0.8000
U60	0.6579	0.6667	0.6905	0.6279	0.6222	0.6250	0.6818	0.6818	0.7500	0.6977	0.7500	0.3659	0.7442	0.7083	0.7143	0.8095	0.8140
U68	0.7429	0.7500	0.7692	0.7250	0.7143	0.7500	0.6977	0.8250	0.7949	0.7805	0.8378	0.3846	0.7619	0.6875	0.7317	0.7857	0.8333
U70	0.7500	0.7222	0.7429	0.7027	0.6842	0.7273	0.6667	0.7838	0.7500	0.7568	0.7941	0.4242	0.6923	0.6591	0.6410	0.6750	0.7250
SCA6	0.7941	0.8421	0.8649	0.8158	0.8000	0.8000	0.7805	0.8500	0.8421	0.8718	0.8649	0.4595	0.8049	0.7727	0.7500	0.8250	0.8750
C42	0.7273	0.7568	0.8286	0.7297	0.7179	0.7353	0.7000	0.7692	0.7838	0.7895	0.8286	0.3243	0.7250	0.7045	0.6923	0.7436	0.7750
C53	0.6923	0.7619	0.7805	0.7209	0.7111	0.7000	0.6957	0.7727	0.7619	0.7907	0.7619	0.4048	0.7556	0.7200	0.6889	0.8140	0.8000
C65	0.7222	0.8158	0.8378	0.7692	0.8000	0.7778	0.7805	0.8049	0.8205	0.8250	0.8158	0.4211	0.8049	0.7234	0.7143	0.7381	0.7674
HTM1	0.7813	0.8571	0.9118	0.8286	0.8108	0.7879	0.7895	0.8158	0.8333	0.8889	0.8824	0.4000	0.8158	0.6889	0.7179	0.7250	0.7750
CCN51	0.7714	0.8205	0.8649	0.7949	0.7805	0.7778	0.7619	0.8293	0.8000	0.8974	0.8421	0.4474	0.7857	0.7447	0.7381	0.7619	0.7907
IMC67	0.7143	0.7895	0.8611	0.7632	0.7949	0.7941	0.8205	0.8000	0.8649	0.8205	0.8611	0.4324	0.8462	0.7333	0.7073	0.7500	0.7619
I12.12	0.8065	0.8529	0.9063	0.8235	0.8056	0.7813	0.8333	0.8108	0.9091	0.8857	0.9091	0.4286	0.8611	0.6977	0.7778	0.8056	0.8158
I14.20	0.7742	0.8000	0.8788	0.7714	0.8056	0.8182	0.8333	0.8611	0.9394	0.8333	0.9091	0.4118	0.8611	0.6818	0.7568	0.7179	0.7500
I16.20	0.7879	0.8611	0.8857	0.8333	0.8649	0.8485	0.8919	0.8684	0.9429	0.8919	0.9143	0.5294	0.9189	0.7500	0.7692	0.7500	0.7805
C69	0.7576	0.7838	0.8571	0.7568	0.7632	0.7576	0.7895	0.7895	0.8571	0.8378	0.8333	0.4167	0.8158	0.6667	0.7368	0.7250	0.7750
C81	0.7143	0.7000	0.7632	0.6750	0.6667	0.6944	0.7073	0.7500	0.8108	0.7692	0.7895	0.3590	0.7561	0.6250	0.7000	0.7143	0.7619
C85	0.8000	0.7714	0.8235	0.7429	0.7297	0.7500	0.7568	0.7838	0.8529	0.8056	0.8788	0.3714	0.7838	0.6591	0.7500	0.7179	0.7692
SilvHu	0.6216	0.7000	0.7632	0.6750	0.6829	0.6857	0.6667	0.7073	0.7179	0.7500	0.7000	0.3784	0.7317	0.7111	0.6585	0.7750	0.7619
H12	0.7059	0.7838	0.8333	0.7568	0.7436	0.7576	0.7692	0.7692	0.8333	0.8378	0.8333	0.3784	0.7949	0.7111	0.7179	0.7949	0.8000
H37	0.4688	0.4571	0.4848	0.5000	0.5000	0.5161	0.5278	0.5429	0.5143	0.5143	0.4857	0.6786	0.5556	0.5122	0.4595	0.5429	0.5405
H61	0.6857	0.7179	0.7778	0.6923	0.6829	0.6857	0.7073	0.7073	0.7632	0.7692	0.7632	0.3947	0.7317	0.6250	0.6585	0.7073	0.7561
H60	0.4286	0.4865	0.4722	0.4595	0.4865	0.4286	0.5405	0.5000	0.4865	0.4737	0.4722	0.8000	0.5263	0.4545	0.4615	0.5000	0.4750
Pand	0.6410	0.7073	0.7692	0.6905	0.6744	0.6757	0.6591	0.6977	0.7073	0.8000	0.7073	0.3500	0.6818	0.7234	0.6512	0.7381	0.7273
PA150	0.6250	0.7143	0.7647	0.6857	0.7222	0.7097	0.7500	0.7297	0.7941	0.7500	0.7879	0.3939	0.7778	0.6512	0.6316	0.6410	0.6923
P7	0.6563	0.6944	0.7429	0.6667	0.6579	0.6364	0.6842	0.6667	0.7429	0.7297	0.7429	0.3056	0.7105	0.6000	0.6316	0.6667	0.7179
ChuC	0.6061	0.6944	0.7143	0.6216	0.6154	0.5882	0.6410	0.6250	0.6757	0.6842	0.6944	0.3429	0.6667	0.5652	0.5750	0.6154	0.6500
ChuS	0.7143	0.7895	0.8158	0.7632	0.7500	0.7429	0.7750	0.7561	0.8158	0.8684	0.8378	0.3947	0.8000	0.7333	0.7500	0.7143	0.7619
SeñoA	0.7097	0.7941	0.8750	0.7647	0.7500	0.7419	0.7297	0.7778	0.7714	0.9091	0.7941	0.4242	0.7568	0.7073	0.7027	0.7027	0.7368
Seño	0.3947	0.4878	0.5000	0.4146	0.4762	0.4615	0.5238	0.5122	0.5000	0.4878	0.4634	0.5625	0.4884	0.4286	0.3778	0.4091	0.4545



## Matriz de Similitud - Parte 2

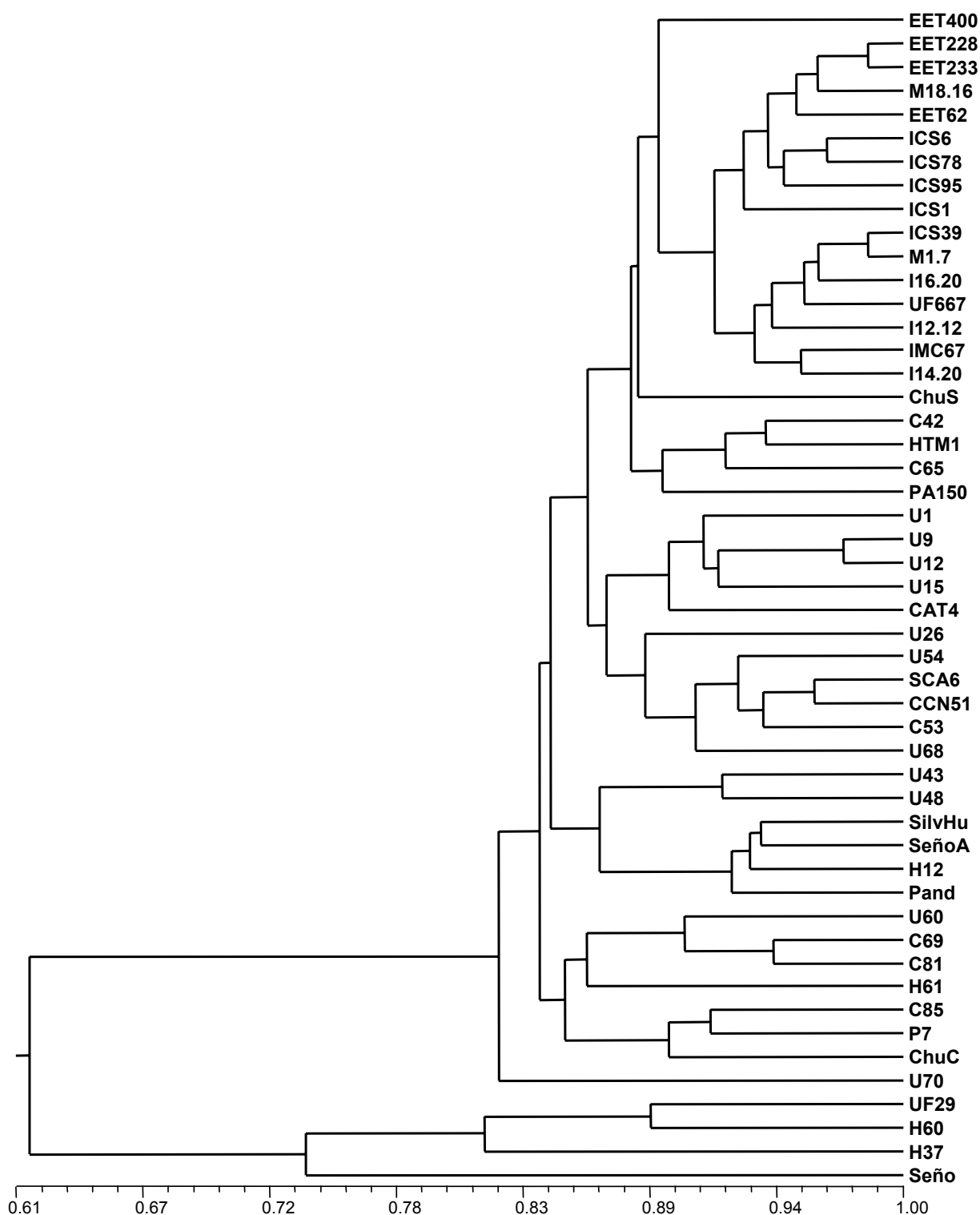
	U15	U26	U43	U48	U54	U60	U68	U70	SCA6	C42	C53	C65	HTM1	CCN51	IMC67	I12.12	I14.20
EET400																	
EET228																	
EET233																	
EET62																	
ICS6																	
ICS78																	
ICS1																	
ICS95																	
ICS39																	
M18.16																	
M1.7																	
UF29																	
UF667																	
U1																	
CAT4																	
U9																	
U12																	
U15	1.000																
U26	0.8333	1.000															
U43	0.7895	0.8000	1.000														
U48	0.7222	0.7568	0.8529	1.000													
U54	0.7609	0.8261	0.7619	0.7895	1.000												
U60	0.7619	0.7333	0.7692	0.7297	0.7660	1.000											
U68	0.7045	0.7500	0.7179	0.7027	0.8000	0.7500	1.000										
U70	0.7179	0.6279	0.6757	0.6471	0.6977	0.6429	0.7895	1.000									
SCA6	0.7619	0.7907	0.7895	0.8235	0.8605	0.7674	0.9000	0.8378	1.000								
C42	0.8649	0.7561	0.7568	0.7879	0.7674	0.7317	0.7750	0.8000	0.8205	1.000							
C53	0.7727	0.8222	0.8049	0.7692	0.8696	0.8222	0.8000	0.7381	0.9048	0.7857	1.000						
C65	0.7442	0.7727	0.7073	0.7105	0.8000	0.7111	0.7273	0.7000	0.8293	0.8421	0.8409	1.000					
HTM1	0.7500	0.7805	0.7568	0.7647	0.7500	0.7143	0.7750	0.7500	0.8684	0.8857	0.7674	0.8684	1.000				
CCN51	0.8095	0.7955	0.7692	0.7105	0.8636	0.6957	0.8333	0.8378	0.9250	0.8462	0.8636	0.8333	0.8462	1.000			
IMC67	0.7805	0.7857	0.7368	0.7222	0.7778	0.7619	0.7619	0.7368	0.8500	0.8649	0.8140	0.8500	0.8649	0.8750	1.000		
I12.12	0.7436	0.8205	0.7222	0.6765	0.7619	0.7500	0.7692	0.6944	0.8649	0.7778	0.8250	0.8158	0.8286	0.8421	0.8611	1.000	
I14.20	0.7895	0.7561	0.6579	0.6286	0.7045	0.7073	0.7250	0.6944	0.7692	0.8286	0.7209	0.8158	0.8286	0.7949	0.9143	0.8788	1.000
I16.20	0.7143	0.7442	0.6923	0.6486	0.6957	0.6977	0.7381	0.7105	0.8250	0.7436	0.7500	0.8250	0.8378	0.8049	0.8684	0.8857	0.8857
C69	0.7073	0.7619	0.7105	0.6857	0.7500	0.8000	0.7143	0.6410	0.8000	0.7179	0.7674	0.7949	0.8108	0.7381	0.8158	0.8857	0.8286
C81	0.7317	0.6889	0.7000	0.6757	0.7021	0.8500	0.7209	0.6500	0.7619	0.7250	0.7556	0.6818	0.7250	0.7045	0.7561	0.7895	0.7632
C85	0.7632	0.6905	0.7027	0.6765	0.6818	0.7692	0.7436	0.7143	0.7895	0.8529	0.6977	0.7436	0.8529	0.7692	0.8333	0.8235	0.8485
SilvHu	0.6744	0.7111	0.7895	0.7714	0.7021	0.7045	0.7209	0.6098	0.7619	0.7250	0.7174	0.6977	0.7692	0.7045	0.7500	0.7368	0.6923
H12	0.7317	0.7674	0.7895	0.7429	0.7333	0.7805	0.7750	0.7027	0.8684	0.8108	0.8095	0.7750	0.8611	0.8000	0.8649	0.8824	0.7838
H37	0.4359	0.3953	0.4167	0.4000	0.4762	0.4750	0.5405	0.5152	0.5405	0.4054	0.4762	0.4615	0.4444	0.5263	0.5135	0.4722	0.4571
H61	0.6364	0.6522	0.6750	0.6667	0.6596	0.6818	0.6977	0.6250	0.7805	0.6585	0.6957	0.6591	0.7000	0.7209	0.7317	0.8108	0.6923
H60	0.3810	0.3864	0.3684	0.3889	0.4545	0.4524	0.4750	0.5000	0.5128	0.3846	0.4545	0.4750	0.4595	0.5000	0.4872	0.4474	0.4324
Pand	0.7209	0.7556	0.7750	0.7297	0.7234	0.6889	0.6889	0.6279	0.7857	0.7317	0.7234	0.6889	0.7750	0.7500	0.7381	0.7436	0.6829
PA150	0.6667	0.6585	0.6667	0.7273	0.6744	0.6750	0.6923	0.6667	0.7838	0.7941	0.6905	0.7838	0.8485	0.7632	0.8824	0.7353	0.7879
P7	0.6500	0.6667	0.7027	0.7059	0.6591	0.7000	0.6750	0.5946	0.7632	0.7222	0.6744	0.6750	0.7714	0.7000	0.7568	0.7941	0.7143
ChuC	0.6500	0.6429	0.6316	0.6857	0.6744	0.6341	0.6098	0.5526	0.6923	0.7222	0.6512	0.6750	0.7222	0.7000	0.7568	0.7647	0.7143
ChuS	0.7805	0.7500	0.7250	0.6579	0.7234	0.6889	0.7209	0.6923	0.8049	0.7500	0.7174	0.7209	0.7949	0.8293	0.8250	0.8108	0.7895
SeñoA	0.7568	0.7692	0.7941	0.7273	0.7143	0.6500	0.6923	0.6571	0.7838	0.7941	0.7073	0.7632	0.8788	0.8108	0.7778	0.7647	0.7647
Seño	0.4000	0.4545	0.3636	0.3846	0.4565	0.3913	0.4348	0.4500	0.4773	0.3721	0.4375	0.4545	0.4286	0.4773	0.4651	0.4500	0.4634

### Matriz de Similitud - Parte 3

	I16.20	C69	C81	C85	SilvHu	H12	H37	H61	H60	Pand	PA150	P7	ChuC	ChuS	SeñoA	Seño
EET400																
EET228																
EET233																
EET62																
ICS6																
ICS78																
ICS1																
ICS95																
ICS39																
M18.16																
M1.7																
UF29																
UF667																
U1																
CAT4																
U9																
U12																
U15																
U26																
U43																
U48																
U54																
U60																
U68																
U70																
SCA6																
C42																
C53																
C65																
HTM1																
CCN51																
IMC67																
I12.12																
I14.20																
I16.20	1.000															
C69	0.8378	1.000														
C81	0.7317	0.8919	1.000													
C85	0.8056	0.8286	0.8333	1.000												
SilvHu	0.7250	0.7500	0.7143	0.7368	1.000											
H12	0.8378	0.8649	0.8205	0.8571	0.8684	1.000										
H37	0.5714	0.5000	0.4737	0.4571	0.5714	0.5000	1.000									
H61	0.7500	0.8158	0.7750	0.7568	0.7750	0.8158	0.5714	1.000								
H60	0.5405	0.4359	0.3810	0.3947	0.4615	0.4103	0.7000	0.4500	1.000							
Pand	0.7143	0.7561	0.7209	0.7436	0.8500	0.8718	0.5000	0.7381	0.4186	1.000						
PA150	0.8000	0.7429	0.6842	0.8125	0.7500	0.8235	0.4706	0.6579	0.4444	0.6923	1.000					
P7	0.7297	0.8000	0.7568	0.8438	0.7105	0.8286	0.3889	0.7778	0.3333	0.7179	0.7813	1.000				
ChuC	0.6842	0.7222	0.6842	0.7879	0.6842	0.7500	0.4054	0.6579	0.3590	0.6500	0.7813	0.8387	1.000			
ChuS	0.8205	0.7750	0.7381	0.8108	0.7143	0.8000	0.4865	0.7500	0.4146	0.7442	0.7568	0.7368	0.6923	1.000		
SeñoA	0.7778	0.8000	0.7297	0.8125	0.8824	0.8788	0.5000	0.7297	0.4054	0.8571	0.8065	0.7273	0.7273	0.8056	1.000	
Seño	0.5122	0.4524	0.4000	0.3810	0.4419	0.4419	0.6364	0.4884	0.5556	0.4444	0.3902	0.3256	0.3415	0.4762	0.4390	1.000

## ANEXO N° 10

Fenograma UPGMA generado con los 50 individuos analizados, utilizando el coeficiente de similitud *Dice*.



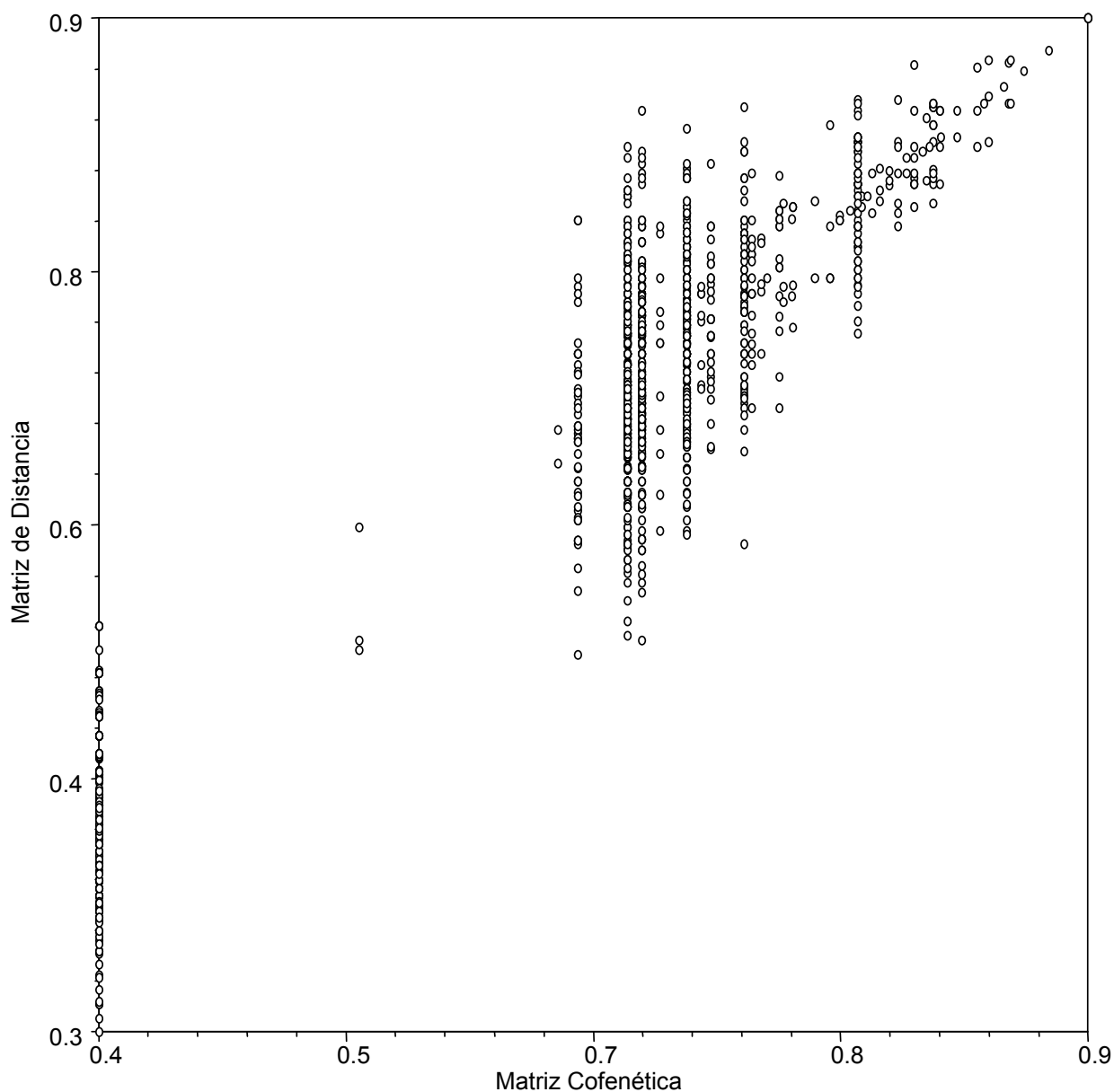
## ANEXO N° 11

Gráficos de la comparación de las dos matrices: cofenética y de distancia. Se presentan para el análisis de los 50 genotipos y para el análisis de los 6 híbridos + 11 parentales. Se muestran los coeficientes de correlación cofenética para cada gráfico.

### Parte 1

50 genotipos

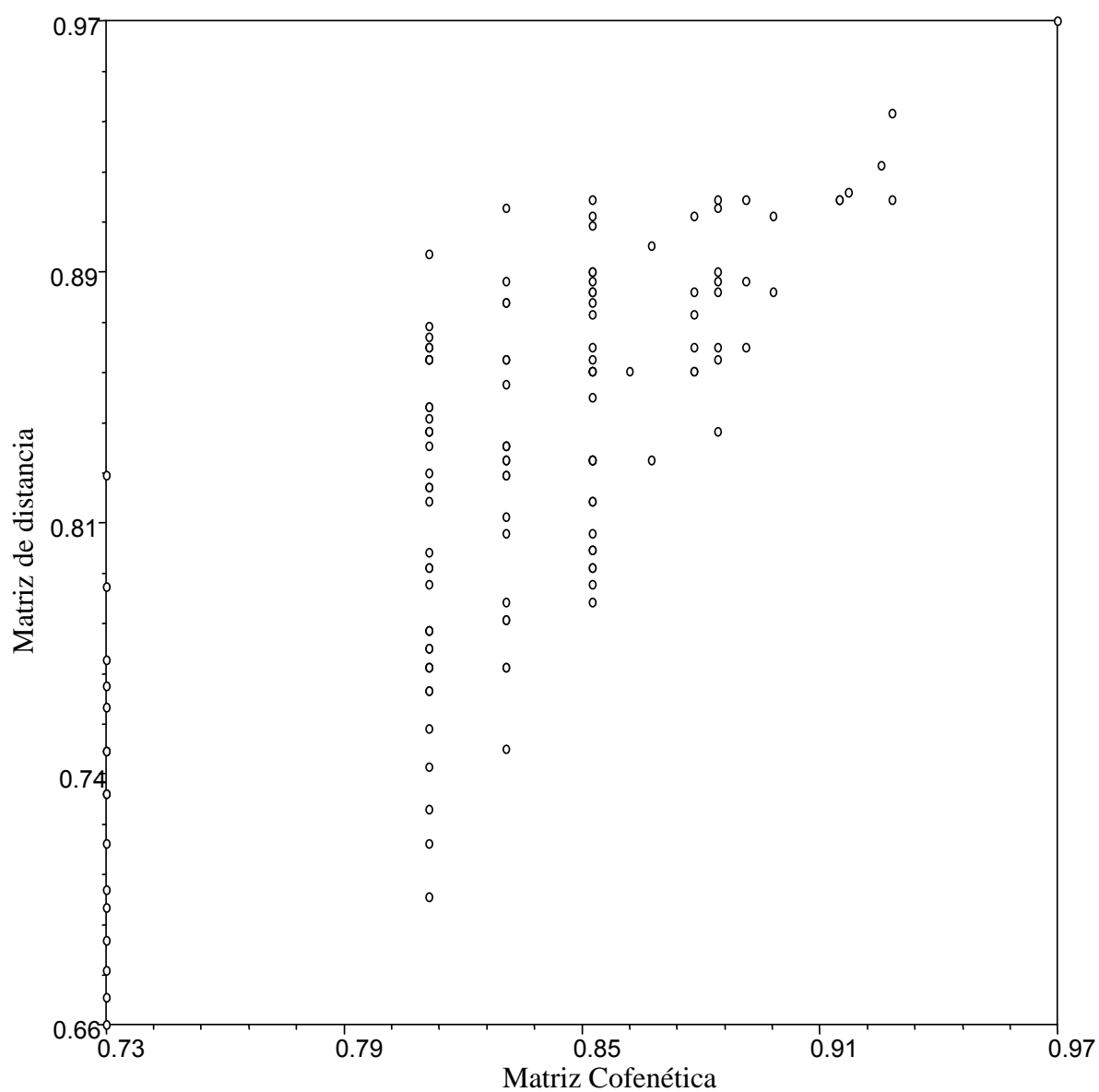
Coefficiente de correlación cofenética = 0.91383



## Parte 2

17 genotipos

Coeficiente de correlación cofenética = 0.77013

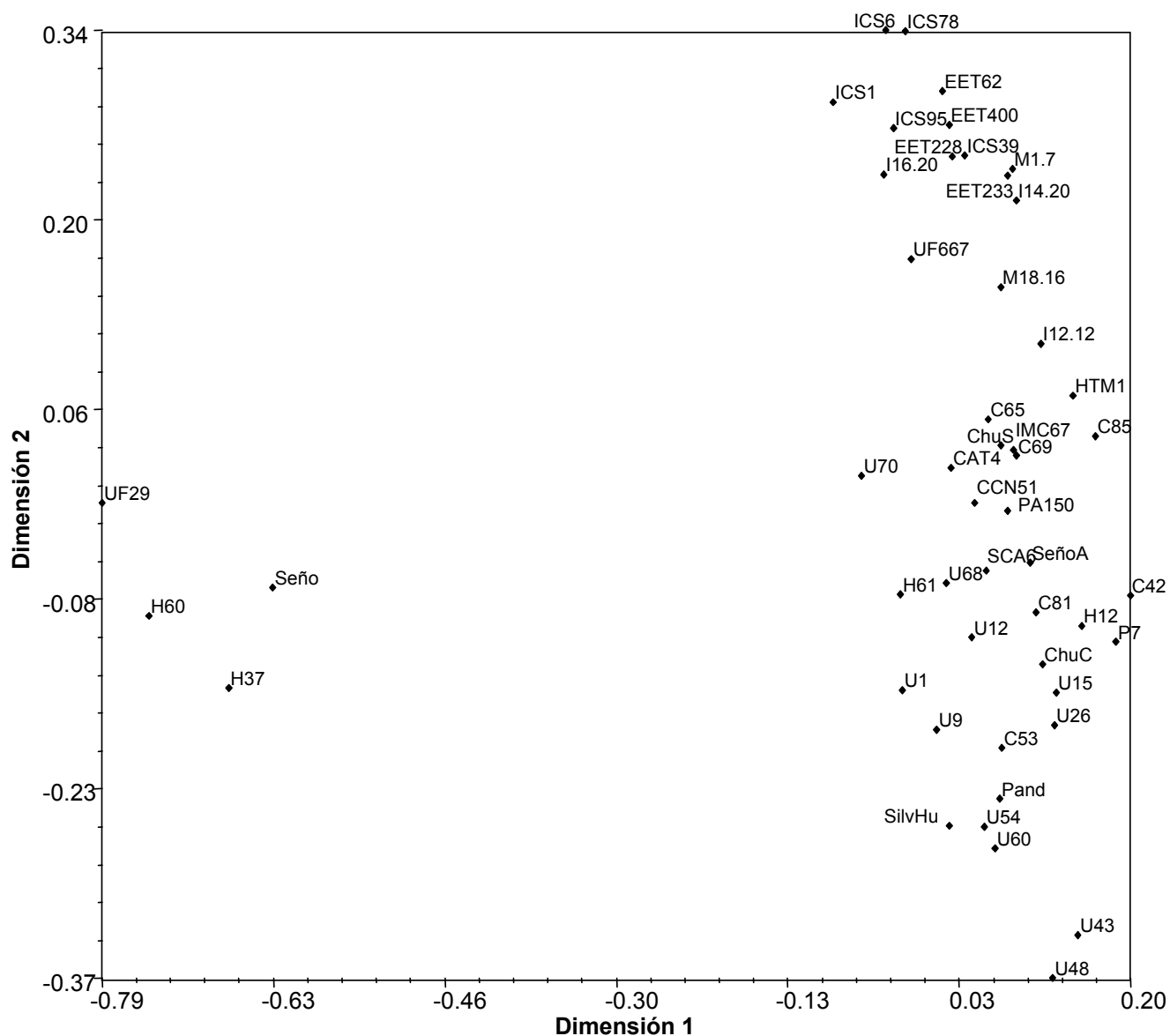


## ANEXO N° 12

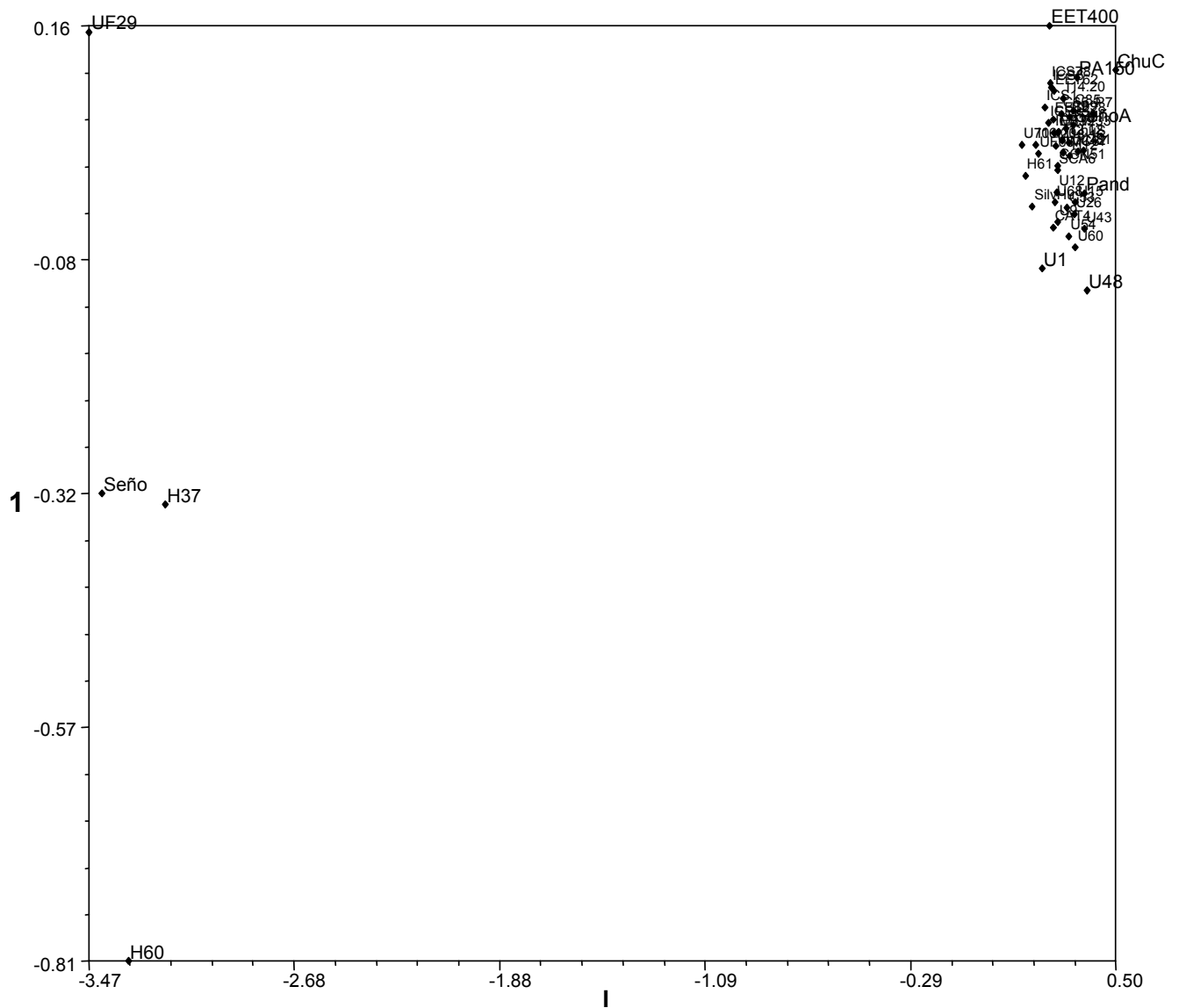
Gráficos de ordenación en dos dimensiones representando a todos los individuos de la colección en Análisis de Coordenadas Principales y Escalamiento Multidimensional no métrico. Se puede observar la gran distancia que hay entre el grupo de 4 accesiones (UF29, H60, H37 y Seño) y el restante. En las Figuras N° 13 y 14 se graficaron el resto de la colección.

### Parte 1

#### Análisis de Coordenadas Principales.



## Escalamiento Multidimensional no métrico.



### Anexo N° 13

Valores G<sub>st</sub> de la comparación par a par entre las subpoblaciones. Entre paréntesis se muestran el número de individuos por población.

		Clones Nacionales						Clones Internacionales				
		Cusco (8)	Huánuco (5)	Junín (1)	Loreto (4)	San Martín (5)	Ucayali (7)	Brasil (1)	Colombia (1)	Ecuador (5)	Trinidad y Tobago (5)	Costa Rica (2)
Clones Nacionales	Cusco (8)	--	0.1393	0.0567	0.1371	0.1087	0.2002	0.0516	0.2154	0.221	0.2757	0.2696
	Huánuco (5)	--	--	0.1761	0.2511	0.2481	0.2777	0.2727	0.2508	0.3097	0.3443	0.1768
	Junín (1)	--	--	--	0.0326	0.0569	0.1269	1	1	-2.6838	-1.7893	0.277
	Loreto (4)	--	--	--	--	0.2008	0.327	0.319	0.3881	0.3754	0.4394	0.4562
	San Martín (5)	--	--	--	--	--	0.1911	-0.01536	0.272	0.2427	0.3206	0.4215
	Ucayali (7)	--	--	--	--	--	--	-0.6661	0.2651	0.2845	0.37847	0.4193
Clones Internacionales	Brasil (1)	--	--	--	--	--	--	--	1	-0.165	0.1473	0.4964
	Colombia (1)	--	--	--	--	--	--	--	--	0.5475	0.6948	0.6123
	Ecuador (5)	--	--	--	--	--	--	--	--	--	0.2755	0.4476
	Trinidad y Tobago (5)	--	--	--	--	--	--	--	--	--	--	0.4188
	Costa Rica (2)	--	--	--	--	--	--	--	--	--	--	--